Statistical Science
2013, Vol. 28, No. 4, 542–563
DOI: 10.1214/13-STS441

Spatial and Spatio-Temporal Log-Gaussian Cox Processes: Extending the Geostatistical Paradigm

Peter J. Diggle, Paula Moraga, Barry Rowlingson and Benjamin M. Taylor

Abstract. In this paper we first describe the class of log-Gaussian Cox processes (LGCPs) as models for spatial and spatio-temporal point process data. We discuss inference, with a particular focus on the computational challenges of likelihood-based inference. We then demonstrate the usefulness of the LGCP by describing four applications: estimating the intensity surface of a spatial point process; investigating spatial segregation in a multi-type process; constructing spatially continuous maps of disease risk from spatially discrete data; and real-time health surveillance. We argue that problems of this kind fit naturally into the realm of geostatistics, which traditionally is defined as the study of spatially continuous processes using spatially discrete observations at a finite number of locations. We suggest that a more useful definition of geostatistics is by the class of scientific problems that it addresses, rather than by particular models or data formats.

Key words and phrases: Cox process, epidemiology, geostatistics, Gaussian process, spatial point process.

1. INTRODUCTION

Spatial statistics has been one of the most fertile areas for the development of statistical methodology during the second half of the twentieth century. A striking, if slightly contrived, illustration of the pace of this development is the contrast between the 90 pages of Bartlett (1975) and the 900 pages of Cressie (1991). Cressie's book established a widely used classification of spatial statistics into three subareas: geostatistical data, lattice data, spatial patterns (meaning point patterns). Within this classification, geostatistical data consist of observed values of some phenomenon of interest associated with a set of spatial locations $x_i : i = 1, \dots, n$, where, in principle, each x_i could have been any location x within a designated spatial region $A \subset \mathbb{R}^2$. Lattice data consist of observed values associated with a fixed set of locations $x_i : i = 1, \dots, n$, that is, the phenomenon of interest exists only at those n specific locations. Finally, in a spatial pattern the data are a set of spatial locations $x_i : i = 1, \dots, n$ presumed to have been generated as a partial realisation of a point process that is itself the object of scientific interest. Almost 20 years later, Gelfand et al. (2010) used the same classification but with a different terminology focused more on the underlying process than on the extant data: continuous spatial variation, discrete spatial variation, and spatial point processes. With this process-based terminology in place, continuous spatial variation implies a stochastic process $\{Y(x) : x \in \mathbb{R}^2\}$, discrete spatial variation implies only a finite-dimensional random variable, $Y = \{Y_i : i = 1, \dots, n\}$, and a point pattern implies a counting measure, $\{dN(x) : x \in \mathbb{R}^2\}$.

Peter J. Diggle is Distinguished University Professor, Lancaster University Medical School, Lancaster, LA1 4YG, United Kingdom and Professor, Institute of Infection and Global Health, University of Liverpool, Liverpool L69 7BE, United Kingdom (e-mail: p.diggle@lancaster.ac.uk). Paula Moraga is Research Associate, Lancaster University Medical School, Lancaster, LA1 4YG, United Kingdom (e-mail: moragase@exchange.lancs.ac.uk). Barry Rowlingson is Research Fellow, Lancaster University Medical School, Lancaster, LA1 4YG, United Kingdom (e-mail: b.rowlingson@lancaster.ac.uk). Benjamin M. Taylor is Lecturer, Lancaster University Medical School, Lancaster, LA1 4YG, United Kingdom (e-mail: b.taylor1@lancaster.ac.uk).

In this paper, we argue first that the most important theoretical distinction within spatial statistics is between spatially continuous and spatially discrete stochastic processes, and second that most natural processes are spatially continuous and should be modelled accordingly. One consequence of this point of view is that in many applications, maintaining a one-to-one linkage between data formats (geostatistical, lattice, point pattern) and associated model classes (spatially continuous, spatially discrete, point process) is inappropriate. In particular, we suggest a redefinition of geostatistics as the collection of statistical models and methods whose purpose is to enable predictive inference about a spatially continuous, incompletely observed phenomenon, $S(x)$, say.

Classically, geostatistical data $Y_i : i = 1, \dots, n$ correspond to noisy versions of $S(x_i)$. A standard geostatistical model, expressed here in hierarchical form, is that $\mathcal{S} = \{S(x) : x \in \mathbb{R}^2\}$ is a Gaussian stochastic process, whilst conditional on $\mathcal{S}$, the Y_i are mutually independent, Normally distributed with means $S(x_i)$ and common variance τ^2. A second scenario, and the focus of the current paper, is when $\mathcal{S}$ determines the intensity, $\lambda(x)$, say, of an observed Poisson point process. An example that we will consider in detail is a log-linear specification, $\lambda(x) = \exp\{S(x)\}$, where $\mathcal{S}$ is a Gaussian process. A third form is when the point process is reduced to observations of the numbers of points Y_i in each of n regions A_i that form a partition (or subset) of the region of interest A. Hence, conditional on $\mathcal{S}$, the Y_i are mutually independent, Poisson-distributed with means

$$\mu_i = \int_{A_i} \lambda(x)\, dx. \tag{1}$$

In the remainder of the paper we show how the log-Gaussian Cox process can be used in a range of applications where $S(x)$ is incompletely observed through the lens of point pattern or aggregated count data. Sections 2 to 4 concern theoretical properties, inference and computation. Section 5 describes several applications. Section 6 discusses the extension to spatio-temporal data. Section 7 gives an outline of how this approach to modelling incompletely observed spatial phenomena extends naturally to the joint analysis of multivariate spatial data when the different data elements are observed at incommensurate spatial scales. Section 8 is a short, concluding discussion.

2. THE LOG-GAUSSIAN COX PROCESS

A (univariate, spatial) Cox process (Cox, 1955) is a point process defined by the following two postulates:

CP1: $\Lambda = \{\Lambda(x) : x \in \mathbb{R}^2\}$ is a nonnegative-valued stochastic process;

CP2: conditional on the realisation $\Lambda(x) = \lambda(x) : x \in \mathbb{R}^2$, the point process is an inhomogeneous Poisson process with intensity $\lambda(x)$.

Cox processes are natural models for point process phenomena that are environmentally driven, much less natural for phenomena driven primarily by interactions amongst the points. Examples of these two situations in an epidemiological context would be the spatial distribution of cases of a noninfectious or infectious disease, respectively. In a noninfectious disease, the observed spatial pattern of cases results from spatial variation in the exposure of susceptible individuals to a combination of observed and unobserved risk-factors. Conditional on exposure, cases occur independently. In contrast, in an infectious disease the observed pattern is at least partially the result of infectious cases transmitting the disease to nearby susceptibles. Notwithstanding this phenomenological distinction, it can be difficult, or even impossible, to distinguish empirically between processes representing stochastically independent variation in a heterogeneous environment and stochastic interactions in a homogeneous environment (Bartlett, 1964).

The moment properties of a Cox process are inherited from those of the process $\Lambda(x)$. For example, in the stationary case the intensity of the Cox process is equal to the expectation of $\Lambda(x)$ and the covariance density of the Cox process is equal to the covariance function of $\Lambda(x)$. Hence, writing $\lambda = \mathrm{E}[\Lambda(x)]$ and $C(u) = \mathrm{Cov}\{\Lambda(x), \Lambda(x-u)\}$, the reduced second moment measure or K-function (Ripley 1976, 1977) of the Cox process is

$$K(u) = \pi u^2 + 2\pi\lambda^{-2} \int_0^u C(v) v\, dv. \tag{2}$$

Møller, Syversveen and Waagepetersen (1998) introduced the class of log-Gaussian processes (LGCPs). As the name implies, an LGCP is a Cox process with $\Lambda(x) = \exp\{S(x)\}$, where $\mathcal{S}$ is a Gaussian process. This construction has an elegant simplicity. One of its attractive features is that the tractability of the multivariate Normal distribution carries over, to some extent, to the associated Cox process.

In the stationary case, let $\mu = \mathrm{E}[S(x)]$ and $C(u) = \sigma^2 r(u) = \mathrm{Cov}\{S(x), S(x-u)\}$. It follows from the moment properties of the log-Normal distribution that the associated LGCP has intensity $\lambda = \exp(\mu + 0.5\sigma^2)$ and covariance density $g(u) = \lambda^2[\exp\{\sigma^2 r(u)\} - 1]$. This makes it both convenient and natural to re-parameterise the model as

$$\Lambda(x) = \exp\{\beta + S(x)\}, \tag{3}$$

where $\mathrm{E}[S(x)] = -0.5\sigma^2$, so that $\mathrm{E}[\exp\{S(x)\}] = 1$ and $\lambda = \exp(\beta)$. This re-parameterisation gives a clean separation between first-order (mean value) and second-order (variation about the mean) properties. Hence, for example, if we wished to model a spatially varying intensity by including one or more spatially indexed explanatory variables $z(x)$, a natural first approach would be to retain the stationarity of $S(x)$ but replace the constant intensity λ by a regression model, $\lambda(x) = \lambda\{z(x); \beta\}$. The resulting Cox process is now an intensity-reweighted stationary point process (Baddeley, Møller and Waagepetersen, 2000), which is the analogue of a real-valued process with a spatially varying mean and a stationary residual.

The definition of a multivariate LGCP is immediate—we simply replace the scalar-valued Gaussian process $S(x)$ by a vector-valued multivariate Gaussian process—and its moment properties are equally tractable. For example, if $S(x)$ is a stationary bivariate Gaussian process with intensities λ_1 and λ_2, and cross-covariance function $C_{12}(u) = \sigma_1\sigma_2 r_{12}(u)$, the cross-covariance density of the associated Cox process is $g_{12}(u) = \lambda_1\lambda_2[\exp\{\sigma_1\sigma_2 r_{12}(u)\} - 1]$.

There is an extensive literature on parametric specifications for the covariance structure of real-valued processes $S(x)$; for a recent summary, see Gneiting and Guttorp (2010a). The theoretical requirement for a function $C(x, y)$ to be a valid covariance function is that it be positive-definite, meaning that for all positive integers n, any associated set of points $x_i \in \mathbb{R}^2 : i = 1, \ldots, n$, and any associated set of real numbers $a_i : i = 1, \ldots, n$,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j C(x_i, x_j) \geq 0. \tag{4}$$

Checking that (4) holds for an arbitrary candidate $C(x, y)$ is not straightforward. In practice, we choose covariance functions from a catalogue of parametric families that are known to be valid. In the stationary case, a widely used family is the Matérn (1960) class $C(u) = \sigma^2 r(u; \phi, \kappa)$, where

$$\begin{aligned} & r(u; \phi, \kappa) \\ & \quad = \{2^{\kappa-1}\Gamma(\kappa)\}^{-1}(u/\phi)^{\kappa} K_{\kappa}(u/\phi) \quad u \geq 0. \end{aligned} \tag{5}$$

In (5), $\Gamma(\cdot)$ is the complete Gamma function, $K_\kappa(\cdot)$ is a modified Bessel function of order κ, and $\phi > 0$ and $\kappa > 0$ are parameters. The parameter ϕ has units of distance, whilst κ is a dimensionless shape parameter that determines the differentiability of the corresponding Gaussian process; specifically, the process is k-times mean square differentiable if $\kappa > k$. This physical interpretation of κ is useful because κ is difficult to estimate empirically (Zhang, 2004), hence, a widely used strategy is to choose between a small set of values corresponding to different degrees of differentiability, for example, $\kappa = 0.5$, 1.5 or 2.5. Estimation of ϕ is more straightforward.

In summary, the LGCP is the natural analogue for point process data of the linear Gaussian model for real-valued geostatistical data (Diggle and Ribeiro, 2007). Like the linear Gaussian model, it lacks any mechanistic interpretation. Its principal virtue is that it provides a flexible and relatively tractable class of empirical models for describing spatially correlated phenomena. This makes it extremely useful in a range of applications where the scientific focus is on spatial prediction rather than on testing mechanistic hypotheses. Section 5 gives several examples.

3. INFERENCE FOR LOG-GAUSSIAN COX PROCESSES

In this section we distinguish between two inferential targets, namely, estimation of model parameters and prediction of the realisations of unobserved stochastic processes. Within the Bayesian paradigm, this distinction is often blurred, because parameters are treated as unobserved random variables and the formal machinery of inference is the same in both cases, consisting of the calculation of the conditional distribution of the target given the data. However, from a scientific perspective parameter estimation and prediction are fundamentally different, because the former concerns properties of the process being modelled whereas the latter concerns properties of a particular *realisation* of that process.

3.1 Parameter Estimation

For parameter estimation, we consider three approaches: moment-based estimation, maximum likelihood estimation, and Bayesian estimation. The first approach is typically very simple to implement and is useful for the initial exploration of candidate models. The second and third are more principled, both being likelihood-based.

3.1.1 *Moment-based estimation.* In the stationary case, moment-based estimation consists of minimising a measure of the discrepancy between empirical and theoretical second-moment properties. One class of such measures is a weighted least squares criterion,

$$D(\theta) = \int_0^{u_0} w(u)\{\hat{K}(u)^c - K(u;\theta)^c\}^2\,du. \tag{6}$$

In the intensity-re-weighted case, (6) can still be used after separately estimating a regression model for a spatially varying $\lambda(x)$ under the working assumption that the data are a partial realisation of an inhomogeneous Poisson process.

This method of estimation has an obviously *ad hoc* quality. In particular, it is difficult to give generally applicable guidance on appropriate choices for the values of u_0 and c in (6). Because the method is intended only to give preliminary estimates, there is something to be said for simply matching $\hat{K}(u)$ and $K(u;\theta)$ by eye. The R (R Core Team, 2013) package lgcp (Taylor et al., 2013) includes an interactive graphics function to facilitate this.

3.1.2 *Maximum likelihood estimation.* The general form of the Cox process likelihood associated with data $X = \{x_i \in A : i = 1, \dots, n\}$ is

$$\begin{aligned}\ell(\theta; X) = \mathrm{P}(X|\theta) &= \int_\Lambda \mathrm{P}(X, \Lambda|\theta)\,d\Lambda \\ &= \mathrm{E}_{\Lambda|\theta}\big(\ell^*(\Lambda; X)\big),\end{aligned} \tag{7}$$

where

$$\ell^*(\Lambda; X) = \prod_{i=1}^n \Lambda(x_i)\left\{\int_A \Lambda(x)\,dx\right\}^{-n} \tag{8}$$

is the likelihood for an inhomogeneous Poisson process with intensity $\Lambda(x)$. The evaluation of (7) involves integration over the infinite-dimensional distribution of Λ. In Section 4.1 below we describe an implementation in which the continuous region of interest A is approximated by a finely spaced regular lattice, hence replacing Λ by a finite set of values $\Lambda(g_k) : k = 1, \dots, N$, where the points $g_1, \dots, g_N$ cover A. Even so, the high dimensionality of the implied integration appears to present a formidable obstacle to analytic progress. One solution, easily stated but hard to implement robustly and efficiently, is to use Monte Carlo methods.

Monte Carlo evaluation of (7) consists of approximating the expectation by an empirical average over simulated realisations of some kind. A crude Monte Carlo method would use the approximation

$$\ell_{\mathrm{MC}}(\theta) = s^{-1}\sum_{j=1}^s \ell\big(\theta; X, \lambda^{(j)}\big), \tag{9}$$

where $\lambda^{(j)} = \{\lambda^{(j)}(g_k) : k = 1, \dots, N\} : j = 1, \dots, s$ are simulated realisations of Λ on the set of grid-points g_k. In practice, this is hopelessly inefficient. A better approach is to use an ingenious method due to Geyer (1999), as follows.

Let $f(X, \Lambda; \theta)$ denote the un-normalised joint density of X and Λ. Then, the associated likelihood is

$$\ell(\theta; X, \Lambda) = f(X, \Lambda; \theta)/a(\theta), \tag{10}$$

where

$$a(\theta) = \int f(X, \Lambda; \theta)\,d\Lambda\,dX \tag{11}$$

is the intractable normalising constant for $f(\cdot)$. It follows that

$$\begin{aligned}&\mathrm{E}_{\theta_0}\big[f(X, \Lambda; \theta)/f(X, \Lambda; \theta_0)\big] \\ &\quad = \iint f(X, \Lambda; \theta)/f(X, \Lambda; \theta_0) \\ &\qquad\quad \times \frac{f(X, \Lambda; \theta_0)}{a(\theta_0)}\,d\Lambda\,dX \\ &\quad = \frac{1}{a(\theta_0)}\int f(X, \Lambda; \theta)\,d\Lambda\,dX \\ &\quad = a(\theta)/a(\theta_0),\end{aligned} \tag{12}$$

where θ_0 is any convenient, fixed value of θ, and E_{θ_0} denotes expectation when $\theta = \theta_0$. However, the function $f(X, \Lambda; \theta)$ in (10) is also an un-normalised conditional density for Λ given X. Under this second interpretation, the corresponding normalised conditional density is $f(X, \Lambda; \theta)/a(\theta|X)$, where

$$a(\theta|X) = \int f(X, \Lambda; \theta)\,d\Lambda, \tag{13}$$

and the same argument as before gives

$$\begin{aligned}&\mathrm{E}_{\theta_0}\big[f(X, \Lambda; \theta)/f(X, \Lambda; \theta_0)|X\big] \\ &\quad = a(\theta|X)/a(\theta_0|X).\end{aligned} \tag{14}$$

It follows from (7), (10) and (13) that the likelihood for the observed data, X, can be written as

$$\ell(\theta; X) = \int \frac{f(x, \Lambda; \theta)}{a(\theta)}\,d\Lambda = a(\theta|X)/a(\theta). \tag{15}$$

Hence, the log-likelihood ratio between any two parameter values, θ and θ_0, is

$$\begin{aligned}&L(\theta; X) - L(\theta_0; X) \\ &\quad = \log\{a(\theta|X)/a(\theta)\} - \log\{a(\theta_0|X)/a(\theta_0)\} \\ &\quad = \log\{a(\theta|X)/a(\theta_0|X)\} - \log\{a(\theta)/a(\theta_0)\}.\end{aligned} \tag{16}$$

Substitution from (12) and (14) gives the result that

$$\begin{aligned} & L(\theta; X) - L(\theta_0; X) \\ & \quad = \log \mathrm{E}_{\theta_0}\left[r(X, \Lambda, \theta, \theta_0) | X\right] \\ & \qquad - \log \mathrm{E}_{\theta_0}\left[r(X, \Lambda, \theta, \theta_0)\right], \end{aligned} \tag{17}$$

where $r(X, \Lambda, \theta, \theta_0) = f(X, \Lambda; \theta)/f(X, \Lambda; \theta_0)$. For any fixed value of θ_0, a Monte Carlo approximation to the log-likelihood, ignoring the constant term $L(\theta_0)$ on the left-hand side of (17), is therefore given by

$$\begin{aligned} \hat{L}(\theta) = & \log\left\{s^{-1}\sum_{j=1}^{s} r\left(X, \lambda^{(j)}, \theta, \theta_0\right)\right\} \\ & - \log\left\{s^{-1}\sum_{j=1}^{s} r\left(X^{(j)}, \lambda^{(j)}, \theta, \theta_0\right)\right\}. \end{aligned} \tag{18}$$

The result (18) provides a Monte Carlo approximation to the log-likelihood function, and therefore to the maximum likelihood estimate $\hat{\theta}$, by simulating the process only at a single value, θ_0. The accuracy of the approximation depends on the number of simulations, s, and on how close θ_0 is to $\hat{\theta}$.

Note that in the second term on the right-hand side of (18) the pairs $(X^{(j)}, \lambda^{(j)})$ are simulated joint realisations of X and Λ at $\theta = \theta_0$, whilst in the first term X is held fixed at the observed data and the simulated realisations $\lambda^{(j)}$ are conditional on X. Conditional simulation of Λ requires Markov chain Monte Carlo (MCMC) methods, for which careful tuning is generally needed. We discuss computational issues, including the design of a suitable MCMC algorithm, in Section 4.

3.1.3 *Bayesian estimation.* One way to implement Bayesian estimation would be directly to combine Monte Carlo evaluation of the likelihood with a prior for θ. However, it turns out to be more efficient to incorporate Bayesian estimation and prediction into a single MCMC algorithm, as described in Section 4.

3.2 Prediction

For prediction, we consider plug-in and Bayesian prediction. Suppose, quite generally, that data Y are to be used to predict a target T under an assumed model with parameters θ. Then, plug-in prediction consists of a series of probability statements within the conditional distribution $[T|Y; \hat{\theta}]$, where $\hat{\theta}$ is a point estimate of θ, whereas Bayesian prediction replaces $[T|Y; \hat{\theta}]$ by

$$[T|Y] = \int [T|Y; \theta][\theta|Y]\, d\theta. \tag{19}$$

This shows that Bayesian prediction is a weighted average of plug-in predictions, with different values of θ weighted according to the Bayesian posterior for θ. The Bayesian solution (19) is the more correct in that it incorporates parameter uncertainty in a way that is both natural, albeit on its own terms, and elegant.

4. COMPUTATION

Inference for LGCPs is a computationally challenging problem. Throughout this section we will use the notation and language of purely spatial processes on $\mathbb{R}^2$, but the discussion applies in more general settings including spatio-temporal LGCPs.

4.1 The Computational Grid

Although we model the latent process S as a spatially continuous process, in practice, we work with a piecewise-constant equivalent to the LGCP model on a collection of cells that form a disjoint partition of the region of interest, A. In the limit as the number of cells tends to infinity, this process behaves like its spatially continuous counterpart. We call the collection of cells on which we represent the process the *computational grid*. The choice of grid reflects a balance between computational complexity and accuracy of approximation. The computational bottleneck arises through the need to invert the covariance matrix, Σ, corresponding to the variance of S evaluated on the computational grid.

Typically, we shall use a computational grid of square cells. This is an example of a *regular grid*, by which we mean that on an extension of the grid notionally wrapped on a torus, a strictly stationary covariance function of the process on $\mathbb{R}^2$ will induce a block-circulant covariance structure on the grid (Wood and Chan, 1994; Møller, Syversveen and Waagepetersen, 1998). For simplicity of presentation, we make no distinction between the extended grid and the original grid, since for extensions that at least double the width and height of the original grid, the toroidal distance between any two cells in the original observation window coincides with their Euclidean distance in $\mathbb{R}^2$. For a second-order stationary process S, inversion of Σ on a regular grid is best achieved using Fourier methods (Frigo and Johnson, 2011). On irregular grids, sparse matrix methods in conjunction with an assumption of low-order Markov dependence are more efficient (Rue and Held, 2005; Rue, Martino and Chopin, 2009; Lindgren, Rue and Lindström, 2011). In this context, Lindgren, Rue and Lindström (2011) demonstrate a link between models assuming a Markov dependence structure and spatially continuous models whose covariance function belongs to a restricted subset of the Matérn class.

4.2 Implementing Bayesian Inference, MCMC or INLA?

We now suppose that the computational grid has been defined and the point process data X have been converted to a set of counts, Y, on the grid cells; note that we envisage using a finely spaced grid, for which cell-counts will typically be 0 or 1. Our goal is to use the data Y to make inferences about the latent process S and the parameters β and θ, which, respectively, parameterise the intensity of the LGCP and the covariance structure of S.

In the Bayesian paradigm we treat S, β and θ as random variables, assign priors to the model parameters (β, θ) and make inferential statements using the posterior/predictive distribution,

$$[S, \beta, \theta|Y] \propto [Y|S, \beta, \theta][S|\theta][\beta, \theta].$$

Two options for computation are as follows: MCMC, which generates random samples from $[S, \beta, \theta|Y]$, and the integrated nested Laplace approximation (INLA), which uses a mathematical approximation.

Taylor and Diggle (2013a) compare the performance of MCMC and INLA for a spatial LGCP with constant expectation β and parameters θ treated as known values. In this restricted scenario, they found that MCMC, run for 100,000 iterations, delivered more accurate estimates of predictive probabilities than INLA. However, they acknowledged that "further research is required in order to design better MCMC algorithms that also provide inference for the parameters of the latent field".

Approximate methods such as INLA have the advantages that they produce results quickly and circumvent the need to assess the convergence and mixing properties of an MCMC algorithm. This makes INLA very convenient for quick comparisons amongst multiple candidate models, which would be a daunting task for MCMC. Against this, MCMC methods are more flexible in that extensions to standard classes of models can usually be accommodated with only a modest amount of coding effort. Also, an important consideration in some applications is that the currently available software implementation of INLA is limited to the evaluation of predictive distributions for univariate, or, at best, low-dimensional, components of the underlying model, whereas MCMC provides direct access to joint posterior/predictive distributions of nonlinear functions of the parameters and of the latent process S. Mixing INLA and MCMC can therefore be a good overall computational strategy. For example, Haran and Tierney (2012) use a heavy-tailed approximation similar in spirit to INLA to construct efficient MCMC proposal schemes.

4.2.1 *Markov Chain Monte Carlo inference for log-Gaussian Cox processes.* MCMC methods generate samples from a Markov chain whose stationary distribution is the target of interest, in our case $[S, \beta, \theta|Y]$. Such samples are inherently dependent but, subject to careful checking of mixing and convergence properties, their empirical distribution is an unbiased estimate of the target, and, in principle, the associated Monte Carlo error can be made arbitrarily small by using a sufficiently long run of the chain. In the current context, we follow Møller, Syversveen and Waagepetersen (1998) and Brix and Diggle (2001) in using a standardised version of S, denoted Γ, and transform θ to the log-scale, so that the MCMC algorithm operates on the whole of $\mathbb{R}^d$, rather than on a restricted subset. We denote the ith sample from the chain by $\zeta^{(i)}$ and write $\pi(\zeta|Y)$ for the target distribution.

The aim in designing MCMC algorithms for any specific class of problems is to achieve faster convergence and better mixing than would be obtained by generic off-the-shelf methods. Gilks, Richardson and Spiegelhalter (1995) and Gamerman and Lopes (2006) give overviews of the extensive literature on this topic. We focus our discussion on the Metropolis-Hastings (MH) algorithm, which includes as a special case the popular Gibbs sampler (Metropolis et al., 1953; Hastings, 1970; Geman and Geman, 1984; Spiegelhalter, Thomas and Best, 1999). In order to use the MH algorithm, we require a proposal density, $q(\cdot|\zeta^{(i-1)})$. At the ith iteration of the algorithm, we sample a candidate, $\zeta^{(i^*)}$, from $q(\cdot)$, and set $\zeta^{(i)} = \zeta^{(i^*)}$ with probability

$$\min\left\{1, \frac{\pi(\zeta^{(i^*)}|Y)}{\pi(\zeta^{(i-1)}|Y)} \frac{q(\zeta^{(i-1)}|\zeta^{(i^*)})}{q(\zeta^{(i^*)}|\zeta^{(i-1)})}\right\},$$

otherwise set $\zeta^{(i)} = \zeta^{(i-1)}$. The choice of $q(\cdot)$ is critical. Previous research on inferential methods for spatial and spatio-temporal log-Gaussian Cox processes has advocated the Metropolis-adjusted Langevin algorithm (MALA), which mimics a Langevin diffusion on the target of interest; see Roberts and Tweedie (1996), Møller, Syversveen and Waagepetersen (1998) and Brix and Diggle (2001); note also Brix and Diggle (2003) and Taylor and Diggle (2013b). Alternatives to

MH include Hamiltonian Monte Carlo methods, as discussed in Girolami and Calderhead (2011).

The Metropolis-adjusted Langevin algorithm exploits gradient information to identify efficient proposals. The algorithms in this article make use of a "pre-conditioning matrix", Ξ (Girolami and Calderhead, 2011), to define the proposal

$$q(\zeta^{(i^*)}|\zeta^{(i-1)}) = \mathrm{N}\Big[\zeta^{(i^*)}; \zeta^{(i-1)} + \frac{h^2}{2}\Xi\nabla\log\{\pi(\zeta^{(i-1)}|Y)\}, h^2\Xi\Big], \tag{20}$$

where h is a scaling constant. Ideally, Ξ should be the negative inverse of the Fisher information matrix evaluated at the maximum likelihood estimate of ζ, that is, $\Xi_{\text{opt}} = \{-\mathrm{E}[\mathcal{I}(\hat{\zeta})]\}^{-1}$ where $\mathcal{I}$ is the observed information. However, this matrix is massive, dense and intractable. In practice, we can obtain an efficient algorithm by choosing Ξ to be an approximation of Ξ_{opt} and further by changing h during the course of the algorithm using adaptive MCMC (Andrieu and Thoms, 2008; Roberts and Rosenthal, 2007). In MALA algorithms, h can be tuned adaptively to achieve an approximately optimal acceptance rate of 0.574 (Roberts and Rosenthal, 2001).

Since the gradient of $\log\pi$ with respect to θ can be both difficult to compute and computationally costly, we instead suggest a random walk proposal for the θ-component of ζ. In the examples described in Section 5 we used the following overall proposal:

$$q(\zeta^{(i^*)}|\zeta^{(i-1)}) = \mathrm{N}\left[\zeta^{(i^*)}; \begin{pmatrix} \Gamma^{(i-1)} + \dfrac{h^2h_\Gamma^2}{2}\Xi_\Gamma\dfrac{\partial\log\{\pi(\zeta^{(i-1)}|Y)\}}{\partial\Gamma} \\ \beta^{(i-1)} + \dfrac{h^2h_\beta^2}{2}\Xi_\beta\dfrac{\partial\log\{\pi(\zeta^{(i-1)}|Y)\}}{\partial\beta} \\ \theta^{(i-1)} \end{pmatrix}, h^2\begin{pmatrix} h_\Gamma^2\Xi_\Gamma & 0 & 0 \\ 0 & h_\beta^2\Xi_\beta & 0 \\ 0 & 0 & ch_\theta^2\Xi_\theta \end{pmatrix}\right]. \tag{21}$$

In (21), Ξ_Γ is an approximation to $\{-\mathrm{E}[\mathcal{I}(\hat{\Gamma})]\}^{-1}$, and similarly for Ξ_β and Ξ_θ. The constants h_Γ^2, h_β^2 and h_θ^2 are the approximately optimal scalings for Gaussian targets explored by the Gaussian random walk or MALA proposals (Roberts and Rosenthal, 2001); these are, respectively, $1.65^2/\dim(\Gamma)^{1/3}$, $1.65^2/\dim(\beta)^{1/3}$ and $2.38^2/\dim(\theta)$, where dim is the dimension.

The acceptance rate for a random walk proposal is often tuned to around 0.234, which is optimal for a Gaussian target in the limit as the dimension of the target goes to infinity. At each step in our algorithm, we jointly propose new values for (S, β) and for θ using, respectively, a MALA and a random walk component in the overall proposal, but we also seek to maintain an acceptance rate of 0.574 to achieve optimality for the MALA parts of the proposal. As a compromise, in our proposal we scale the matrix Ξ_θ by a constant factor c and the proposal covariance matrix by a single adaptive h. In the examples described in Section 5 we used a value of $c = 0.4$, which appears to work well across a range of scenarios.

5. APPLICATIONS

5.1 Smoothing a Spatial Point Pattern

The intensity, $\lambda(x)$, of an inhomogeneous spatial point process is the unique nonnegative valued function such that the expected number of points of the process, called *events*, that fall within any spatial region B is

$$\mu(B) = \int_B \lambda(x)\,dx. \tag{22}$$

Suppose that we wish to estimate $\lambda(x)$ from a partial realisation consisting of all of the events of the process that fall within a region A, hence, $X = \{x_i \in A : i = 1, \ldots, n\}$. Figure 1 shows an example in which the data are the locations of 703 hickory trees in a 19.6 acre (281.6 by 281.6 metre) square region A (Gerrard, 1969), which we have re-scaled to be of dimension 100 by 100.

An intuitively reasonable class of estimators for $\lambda(x)$ is obtained by counting the number of events that lie within some fixed distance, h, say, of x and dividing by πh^2 or, to allow for edge-effects, by the area, $B(x, t)$, of the intersection of A and a circular disc with centre x and radius h, hence,

$$\tilde{\lambda}(x; h) = B(x; h)^{-1}\sum_{i=1}^{n} I(\|x - x_i\| \le h). \tag{23}$$

This estimate is, in essence, a simple form of bivariate kernel smoothing with a uniform kernel function

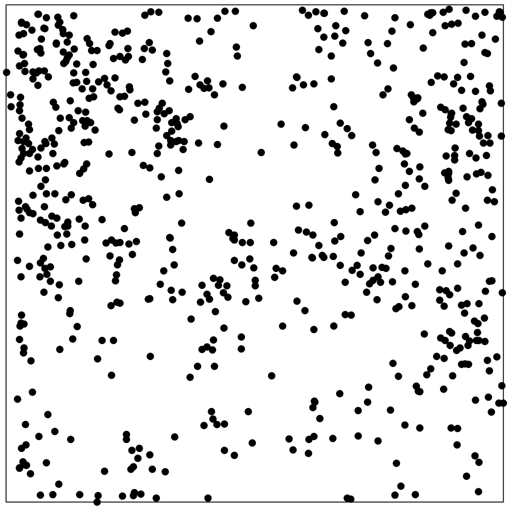

FIG. 1. *Locations of* 703 *hickories in a* 19.6 *acre square plot, re-scaled to* 100 *by* 100 *units* (Gerrard, 1969).

(Silverman, 1986). Berman and Diggle (1989) derived the mean square error of (23) as a function of h under the assumption that the underlying point process is a stationary Cox process. They then showed how to estimate, and thereby approximately minimise, the mean square error without further parametric assumptions.

A different way to formalise the smoothing problem is as a prediction problem associated with the log-Gaussian Cox process, (3). In this formulation, $\Lambda(x) = \exp\{\beta + S(x)\}$, where $S(\cdot)$ is a stationary Gaussian process indexed by a parameter θ and the target for prediction is $\Lambda(x)$. The formal solution is the predictive distribution of $\Lambda(\cdot)$ given X. For a smooth estimate, analogous to (23), we take $\hat{\lambda}(x)$ to be a suitable summary of the predictive distribution, for example, its point-wise expectation or median. This is still a nonparametric solution, in the sense that no parametric form is specified in advance for $\hat{\lambda}(x)$. The parameterisation of the Gaussian process $S(\cdot)$ is the counterpart of the choices made in the kernel estimation approach, namely, the specification of the uniform kernel in (23) and the value of the bandwidth, h.

For this application, we specify that $S(\cdot)$ has mean $-0.5\sigma^2$, variance σ^2 and exponential correlation function, $r(u) = \exp(-u/\phi)$, hence, $\theta = (\sigma^2, \phi)$. We conduct Bayesian predictive inference using MCMC methods implemented in an extension of the `R` package `lgcp` (Taylor et al., 2013). For β we chose a diffuse prior, $\beta \sim N(0, 10^6)$. For σ and ϕ, we chose Normal priors on the log scale: $\log\sigma \sim N(\log(1), 0.15)$ and $\log\phi \sim N(\log(10), 0.15)$. We initialised the MCMC as follows. For σ and ϕ, we minimised

$$\int_0^{25} \left(\hat{K}(r)^{0.25} - K(r;\sigma,\phi)^{0.25}\right)^2 dr,$$

where $K(r;\sigma,\phi)$ is the K-function of the model and $\hat{K}(r)$ is Ripley's estimate (Ripley 1976, 1977), resulting in initial values of $\sigma = 0.50$ and $\phi = 12.66$. The initial value of Γ was set to a 256×256 matrix of zeros and β was initialised using estimates from an overdispersed Poisson generalised linear model fitted to the cell counts, ignoring spatial correlation.

For the MCMC, we used a burn-in of 100,000 iterations followed by a further 900,000 iterations, of which we retained every 900th iteration so as to give a weakly dependent sample of size 1000. Convergence and mixing diagnostics are shown in the supplementary material [Diggle et al. (2013)]. Figure 2 compares the prior

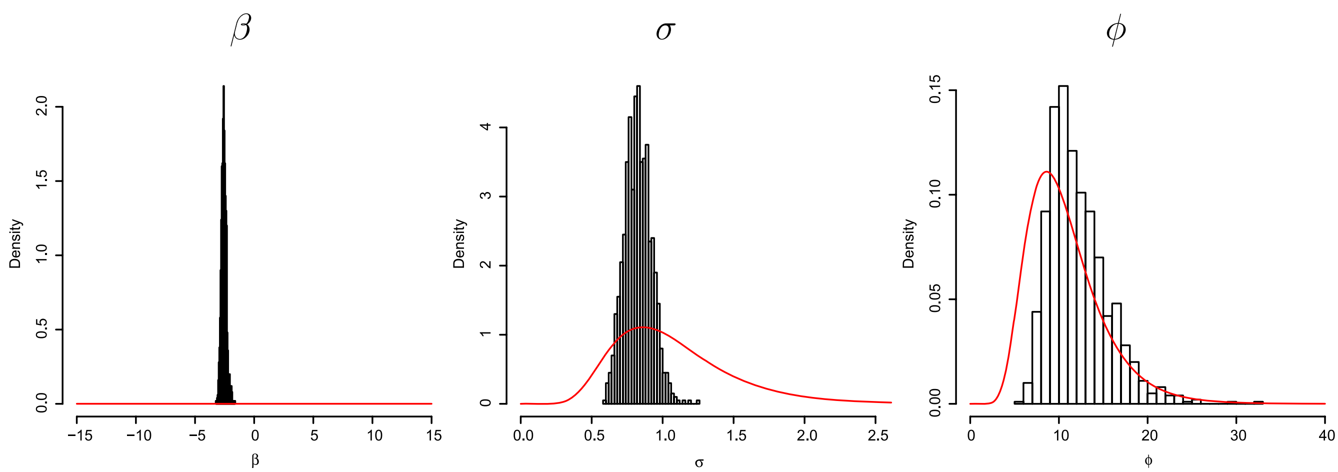

FIG. 2. *Prior* (*continuous curve*) *and posterior* (*histogram*) *distributions for the parameters* β, σ *and* ϕ *in the LGCP model for the hickory data.*

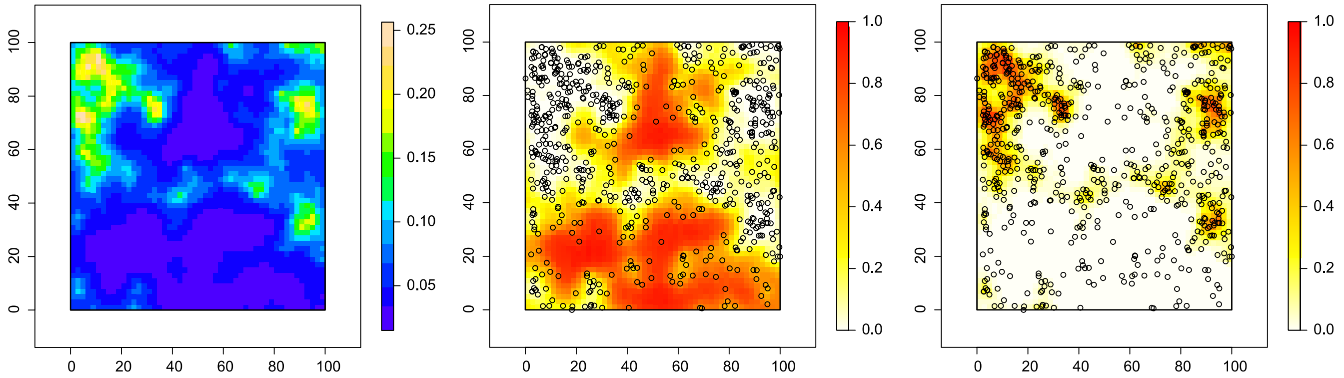

FIG. 3. *Left*: 50% *posterior percentiles of* $\Lambda(x) = \exp\{\beta + S(x)\}$ *for the hickory data. Middle*: *plot of posterior* $\mathbb{P}\{\exp[S(x)] < 1/2\}$. *Right*: *plot of posterior* $\mathbb{P}\{\exp[S(x)] > 2\}$. *Middle and right plots also show the locations of the trees.*

and posterior distributions of the three model parameters showing, in particular, that the data give only weak information about the correlation range parameter, ϕ. This is well known in the classical geostatistical context where the data are measured values of $S(x)$ (see, e.g., Zhang, 2004), and is exacerbated in the point process setting.

The left plot in Figure 3 shows the pointwise 50th percentiles of the predictive distribution for the target, $\Lambda(x)$ over the observation window; this clearly identifies the pattern of the spatial variation in the intensity. The LGCP-based solution also enables us to map areas of particularly low or high intensity. The middle and right plots in Figure 3 are maps of $\mathbb{P}\{\exp[S(x)] < 1/2\}$ and $\mathbb{P}\{\exp[S(x)] > 2\}$. The areas in these plots where the posterior probabilities are high correspond, respectively, to areas where the density of trees is less than half and more than double the mean density.

The LGCP-based solution to the smoothing problem is arguably over-elaborate by comparison with simpler methods such as kernel smoothing. Against this, arguments in its favour are that it provides a principled rather than an *ad hoc* solution, probabilistic prediction rather than point prediction, and an obvious extension to smoothing in the presence of explanatory variables by specifying $\Lambda(x) = \exp\{u(x)'\beta + S(x)\}$, where $u(x)$ is a vector of spatially referenced explanatory variables.

5.2 Spatial Segregation: Genotypic Diversity of Bovine Tuberculosis in Cornwall, UK

Our second application concerns a multivariate version of the smoothing problem described in Section 5.1. Events are now of k types, hence, the data are $X = \{X_j : j = 1, \dots, k\}$, where $X_j = \{x_{ij} \in A : i = 1, \dots, n_j\}$ and the corresponding intensity functions are $\lambda_j(x) : j = 1, \dots, k$. Write $\lambda(x) = \sum_{j=1}^{k} \lambda_j(x)$ for the intensity of the superposition. Under the additional assumption that the underlying process is an inhomogeneous Poisson process, then conditional on the superposition, the labellings of the events are a sequence of independent multinomial trials with position-dependent multinomial probabilities,

$$\begin{aligned} p_j(x) &= \lambda_j(x)/\lambda(x) \\ &= \mathrm{P}(\text{event at location } x \text{ is of type } j) \\ &\qquad j = 1, \dots, k. \end{aligned}$$

A basic question for any multivariate point process data is whether the type-specific component processes are independent. When they are not, further questions of interest are context-specific. Here, we describe an analysis of data relating to bovine tuberculosis in the county of Cornwall, UK.

Bovine tuberculosis (BTB) is a serious disease of cattle. It is endemic in parts of the UK. As part of the national control strategy, herds are regularly inspected for BTB. When disease in a herd is detected and at least one tuberculosis bacterium is successfully cultured, the genotype that is responsible for the BTB breakdown can be determined. Here, we re-visit an example from Diggle, Zheng and Durr (2005) in which the events are the locations of cattle herds in the county of Cornwall, UK, that have tested positive for bovine BTB over the period 1989 to 2002, labelled according to their genotypes. The data, shown in Figure 4, are limited to the 873 locations with the four most common genotypes; six less common genotypes accounted for an additional 46 cases.

The question of primary interest in this example is whether the genotypes are randomly intermingled

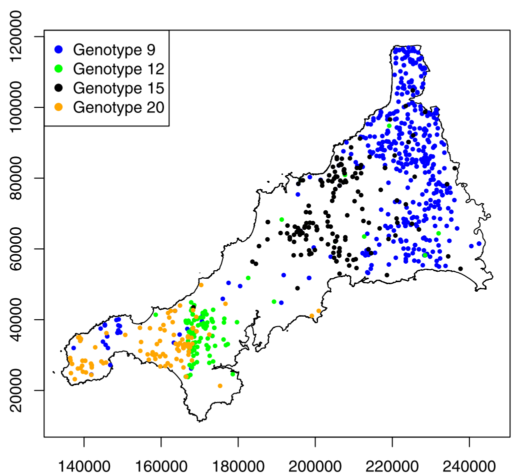

FIG. 4. *Locations of cattle herds in Cornwall, UK, that have tested positive for bovine tuberculosis (BTB) over the period* 1989 *to* 2002. *Points are coded according to the genotype of the infecting BTB organism.*

amongst the locations and, if not, to what extent specific genotypes are spatially segregated. This question is of interest because the former would be consistent with the major transmission mechanism being cross-infection during the county-wide movement of animals to and from markets, whereas the latter would be indicative of local pools of infection, possibly involving transmission between cattle and reservoirs of infection in local wildlife populations (Woodroffe et al., 2005; Donnelly et al., 2006).

To model the data, we consider a multivariate log-Gaussian Cox process with

$$\Lambda_k(x) = \exp(\beta_k + S_0(x) + S_k(x)) \qquad k = 1, \ldots, m. \tag{24}$$

In (24), $m = 4$ is the number of genotypes, the parameters β_k relate to the intensities of the component processes, $S_0(x)$ is a Gaussian process common to all types of points and the $S_k(x) : k = 1, \ldots, m$ are Gaussian processes specific to each genotype. Although $S_0(x)$ is not identifiable from our data without additional assumptions, its inclusion helps the interpretation of the model, in particular, by emphasising that the component intensities $\Lambda_k(x)$ are not mutually independent processes.

In this example, we used informative priors for the model parameters: $\log \sigma \sim \mathrm{N}(\log 1.5, 0.015)$, $\log \phi \sim \mathrm{N}(\log 15{,}000, 0.015)$ and $\beta_k \sim \mathrm{N}(0, 10^6)$. Because the algorithm mixes slowly, this proved to be a very challenging computational problem. For the MCMC, we used a burn-in of 100,000 iterations followed by a further 18,000,000 iterations, of which we retained every 18,000th iteration so as to give a sample of size 1000. Convergence, mixing diagnostics and plots of the prior and posterior distributions of σ and ϕ are shown in the supplementary material [Diggle et al. (2013)]. These plots show that the chain appeared to have reached stationarity with low autocorrelation in the thinned output. The plots also illustrate that there is little information in the data on σ and ϕ.

Within (24) the hypothesis of randomly intermingled genotypes corresponds to $S_k(x) = 0 : k = 1, \ldots, 4$, for all x. Were it the case that farms were uniformly distributed over Cornwall, $S_0(x)$ would then represent the spatial variation in the overall risk of BTB, irrespective of genotype. Otherwise, $S_0(x)$ conflates spatial variation in overall risk with the spatial distribution of farms. For the Cornwall BTB data the evidence against randomly intermingled genotypes is overwhelming and we focus our attention on spatial variation in the probability that a case at location x is of type k, for each of $k = 1, \ldots, 4$. These conditional probabilities are

$$p_k(x) = \frac{\Lambda_k(x)}{\sum_{j=1}^m \Lambda_j(x)} = \exp\Bigl[-\sum_{j \neq k}\{\beta_j + S_j(x)\}\Bigr]$$

and do not depend on the unidentifiable common component $S_0(x)$. Figure 5 shows point predictions of the four genotype-specific probability surfaces, defined as the conditional expectations $\hat{p}_k(x) = \mathrm{E}[p_k(x)|X]$ for each of $k = 1, \ldots, 4$.

As argued earlier, one advantage of a model-based approach to spatial smoothing is that results can be presented in ways that acknowledge the uncertainty on the point predictions. We could replace each panel of Figure 5 by a set of percentile plots, as in Figure 3. For an alternative display that focuses more directly on the core issue of spatial segregation, let $A_k(c, q)$ denote the set of locations x for which $\mathrm{P}\{p_k(x) > c|X\} > q$. As c and q both approach 1, each $A_k(c, p)$ shrinks towards the empty set, but more slowly in a highly segregated pattern than in a weakly segregated one. In Figure 6 we show the areas $A_k(0.8, q)$ for each of $q = 0.6, 0.7, 0.8$ and 0.9. Genotype 9, which contributes 494 to the total of 873 cases, dominates strongly in an area to the east and less strongly in a smaller area to the west. Genotype 15 contributes 166 cases and dominates in a single, central area. Genotypes 12 and 20 each contribute a proportion of approximately 0.12 to the total, with only small pockets of dominance to the south-west.

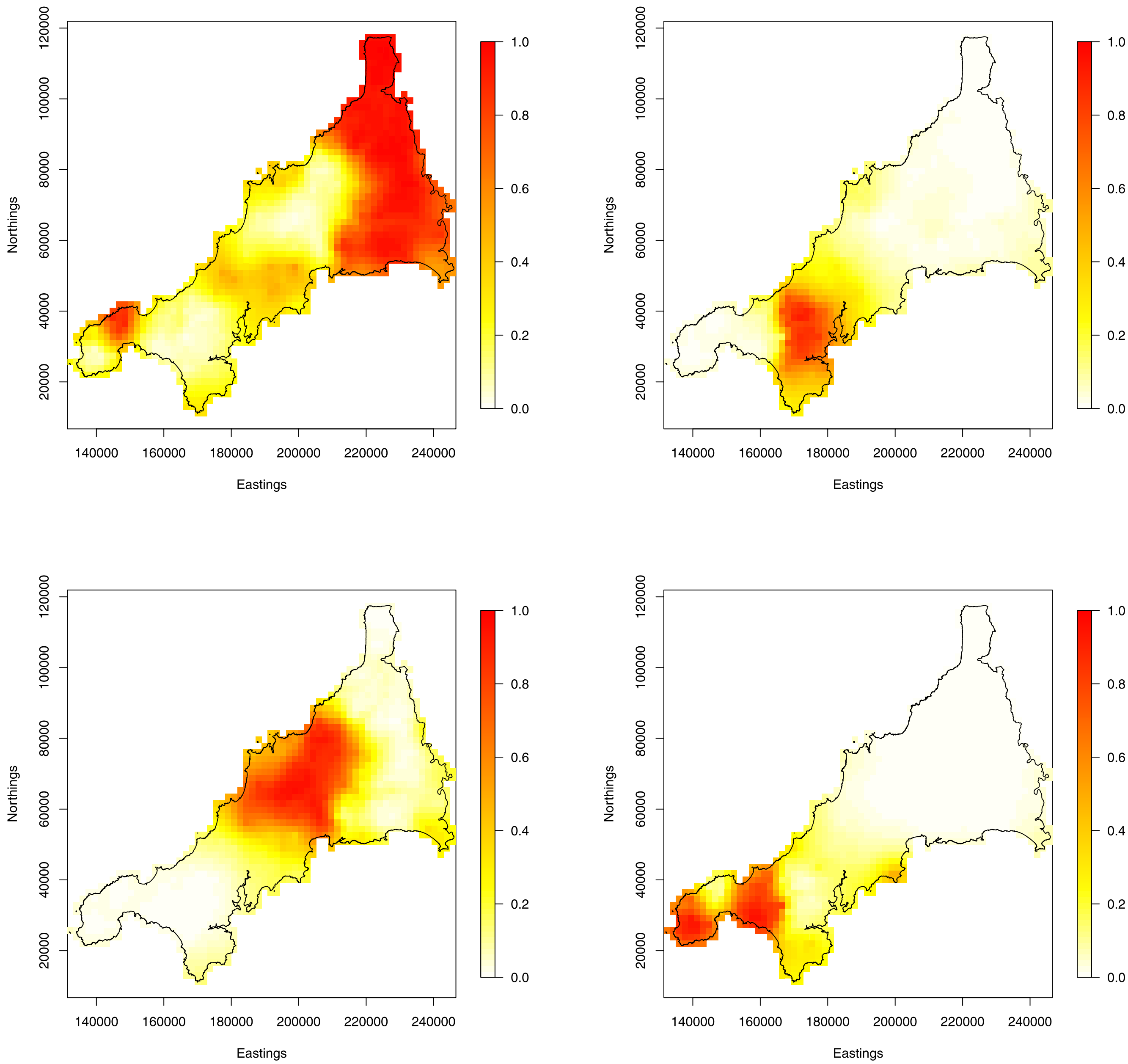

FIG. 5. *Genotype-specific probability surfaces for the Cornwall BTB data. Upper-left panel corresponds to genotype* 9, *upper-right to genotype* 12, *lower-left to genotype* 15, *lower-right to genotype* 20.

If infection times were known, we could perform inference via MCMC under a spatio-temporal version of the model,

$$\Lambda_k(x,t) = \exp\bigl(Z_k(x,t)\beta_k + S_0(x,t) + S_k(x,t)\bigr)$$
$$k = 1, \dots, m,$$

with $\Lambda_k(x,t)$, and $S_k(x,t)$ for $k = 0, \dots, m$ spatio-temporal versions of the purely spatial processes in (24) and $Z_k(x,t)$ a vector of spatio-temporal covariates. Unlike purely spatial models, spatio-temporal models are potentially able to investigate mechanistic hypotheses about disease transmission. For example, in the context of this example a spatio-temporal analysis could distinguish between segregated patches that are stable over time or that grow from initially isolated cases.

5.3 Disease Atlases

Figure 7 is a typical example of the kind of map that appears in a variety of cancer atlases. This ex-

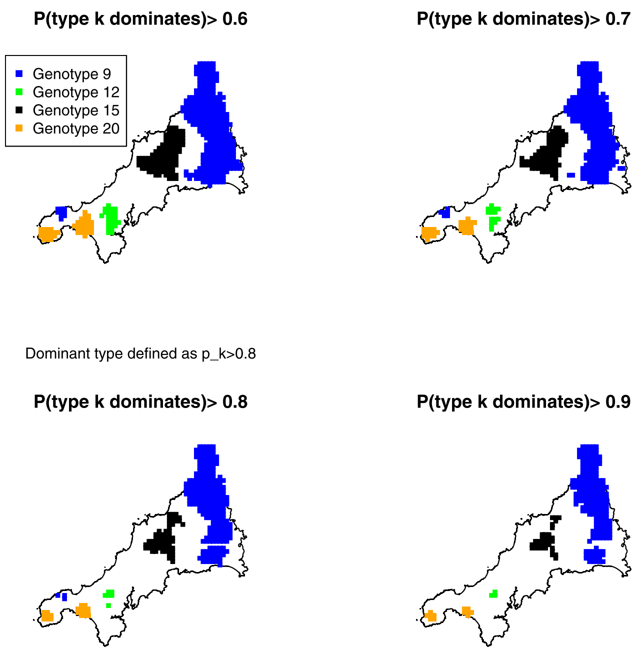

FIG. 6. *k-dominant areas for each of the four genotypes in the Cornwall data.*

ample is taken from a Spanish national disease atlas project (López-Abente et al., 2006). The map estimates the spatial variation in the relative risk of lung cancer in the Castile-La Mancha Region of Spain and some surrounding areas. It is of a type known to geographers as a choropleth map, in which the geographical region of interest, A, is partitioned into a set of subregions A_i and each subregion is colour-coded according to the numerical value of the quantity of interest. The standard statistical methodology used to convert data on case-counts and the number of people at risk in each subregion is the following hierarchical Poisson-Gaussian Markov random field model, due to Besag, York and Molié (1991).

Let Y_i denote the number of cases in subregion A_i and E_i a standardised expectation computed as the expected number of cases, taking into account the demographics of the population in subregion A_i but assuming that risk is otherwise spatially homogeneous. Assume that the Y_i are conditionally independent Poisson-distributed conditional on a latent random vector $S = (S_1, \ldots, S_m)$, with conditional means $\mu_i = E_i \exp(\alpha + S_i)$. Finally, assume that S is multivariate Gaussian, with its distribution specified as a Gaussian Markov random field (Rue and Held, 2005). A Markov random field is a multivariate distribution specified indirectly by its full conditionals, $[S_i | S_j : j \neq i]$. In the Besag, York and Molié (1991) model the full conditionals take the so-called *intrinsic autoregressive* form,

$$S_i | S_j : j \neq i \sim \mathrm{N}(\bar{S}_i, \tau^2/n_i), \tag{25}$$

where $\bar{S}_i = n_i^{-1} \sum_{j \sim i} S_j$ is the mean of the S_j over subregions A_j considered to be *neighbours* of A_i and n_i is the number of such neighbours. Typically, subregions are defined to be neighbours if they share a common boundary.

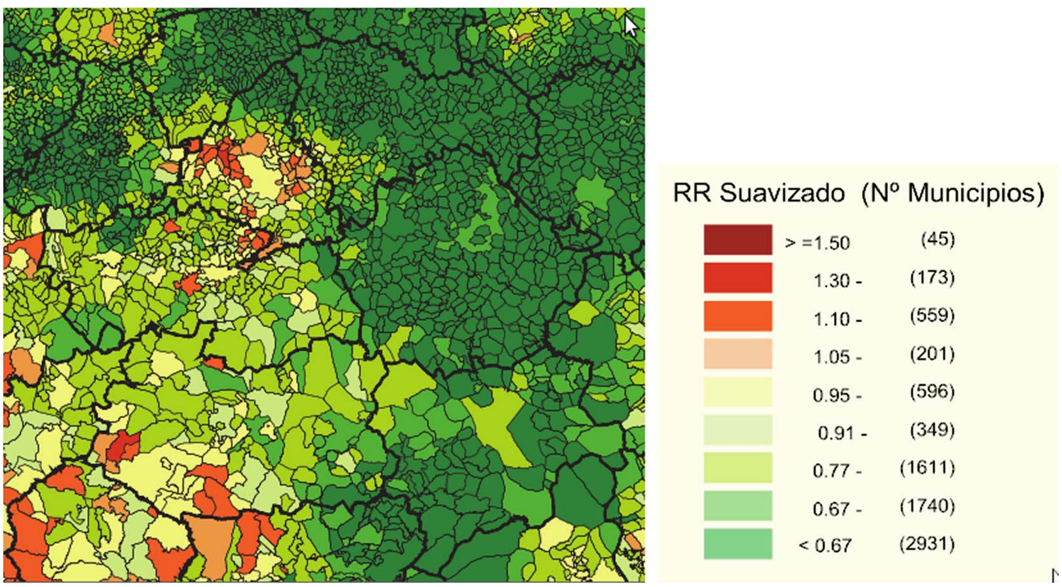

FIG. 7. *Lung Cancer mortality in the Castile-La Mancha Region of Spain. Figure reproduced from page* 42 *of López-Abente et al.* (2006) *by kind permission of the authors.*

An alternative approach is to model the locations of individual cancer cases as an LGCP with intensity $\Lambda(x) = d(x)R(x)$, where $d(x)$ represents population density, assumed known, and $R(x)$ denotes disease risk, $R(x) = \exp\{S(x)\}$. Conditional on $R(\cdot)$, case-counts in subregions A_i are independent and Poisson-distributed with means

$$\mu_i = \int_{A_i} d(x)R(x)\,dx.$$

This approach leads to spatially smooth risk-maps whose interpretation is independent of the particular partition of A into subregions A_i. This is an important consideration when the A_i differ greatly in size and shape, as the definition of neighbours in an MRF model then becomes problematic; see, for example, Wall (2004). Fitting a spatially continuous model also has the potential to add information to an analysis of aggregated data, for example, when data on environmental risk-factors are available at high spatial resolution. A caveat is that the population density may only be available in the form of small-area population counts, implying a piece-wise constant surface $d(x)$ that can only be a convenient fiction. Note, however, that spatially continuous modelled population density maps have been constructed and are freely available; see, for example, http://sedac.ciesin.columbia.edu/data/set/gpw-v3-population-density.

For the Spanish lung cancer data, we have covariate information available at small-area, which we incorporate by fitting the model

$$\Lambda(x) = d(x)\exp\{z(x)'\beta + S(x)\}, \tag{26}$$

treating the covariate surfaces $z(x)$ as piece-wise constant.

For Bayesian inference under the continuous model (26) we follow Li et al. (2012) by adding standard data augmentation techniques to the MCMC fitting algorithm described earlier. Recall that for computational purposes, we perform all calculations on a fine grid, treating the cell counts in each grid cell as Poisson distributed conditional on the latent process $S(\cdot)$. Provided the computational grid is fine enough, each A_i can be approximated by the union of a set of grid cells, and we can use a grid-based Gibbs sampling strategy, repeatedly sampling first from $[S, \beta, \theta | N, Y_+] = [S, \beta, \theta | N]$ and then from $[N | S, \beta, \theta, Y_+]$, where N are the cell counts on the computational grid, $Y_+ = \{Y_i = \sum_{x \in A_i} N(x) : i = 1, \ldots, m\}$ and θ parameterises the covariance structure of S. Sampling from the first of these densities can be achieved using a Metropolis-Hastings update as discussed in Section 4. The second density is a multinomial distribution and poses no difficulty.

Our priors for this example were as follows: $\log\sigma \sim \mathrm{N}(\log 1, 0.3)$, $\log\phi \sim \mathrm{N}(\log 3000, 0.15)$ and $\beta \sim \mathrm{MVN}(0, 10^6 I)$. For the MCMC algorithm, we used a burn-in of 100,000 iterations followed by a further 18,000,000 iterations, of which we retained every 18,000th iteration so as to give a sample of size 1000. Convergence, mixing diagnostics and plots of the prior and posterior distributions of σ and ϕ are shown in the supplementary material [Diggle et al. (2013)]. As in the Cornwall BTB analysis, these plots indicated convergence to the stationary distribution and low autocorrelation in the thinned output.

In the analysis reported here, we base our offset on modelled population data at 100 metre resolution obtained from the European Environment Agency; see http://www.eea.europa.eu/data-and-maps/data/population-density-disaggregated-with-corine-land-cover-2000-2. We projected this very fine population information onto our computational grid, which consisted of cells 3100 × 3100 metres in dimension. We used an exponential model for the covariance function of $S(\cdot)$ and estimated its parameters (posterior median and 95% credible interval) to be $\sigma = 1.57$ (1.45, 1.71) and $\phi = 1294$ (814, 1849) metres. Figure 8 illustrates the shape of the posterior covariance function; it can be seen from this plot that the posterior dependence between cells is over a relatively small range.

Table 1 summarises our estimation of covariate effects. Our results show that estimated (posterior median) mortality rates were higher in areas with higher

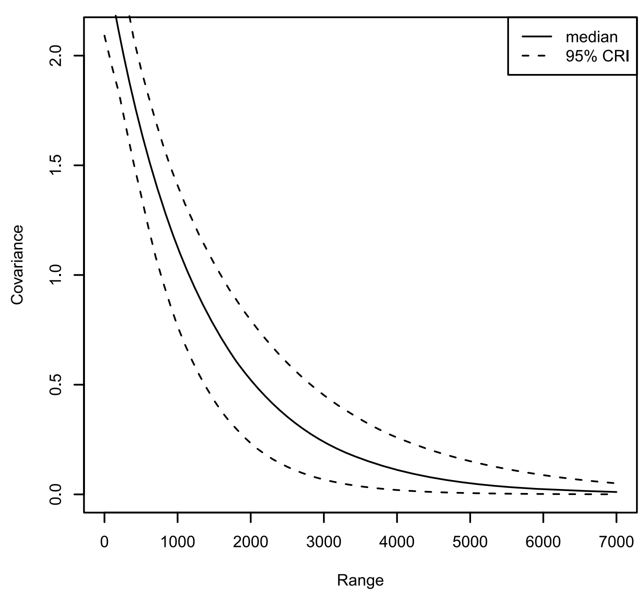

FIG. 8. *Posterior covariance function.*

TABLE 1
Selected quantiles of the posterior distributions of standardised covariate effects for the Spanish lung cancer data

	Quantile		
Parameter	**0.50**	**0.025**	**0.975**
Percentage illiterate	1.13	1.03	1.24
Percentage unemployed	0.92	0.8	1.03
Percentage farmers	0.88	0.76	1.00
Percentage of people over 65 years old	1.2	0.96	1.51
Income index	1.19	1.03	1.39
Average number of people per home	0.98	0.75	1.26

rates of illiteracy and higher income; these effects were statistically significant at the 5% level, in the sense that the Bayesian 95% credible intervals excluded zero. The remaining covariates (unemployment, percentage farmers, percentage of people over 65 and average number of people per home) had a protective effect, but only significantly so in the case of percentage farmers.

Figure 9 shows the resulting maps. The top left-hand panel shows the predicted, covariate-adjusted relative risk surface derived from the log-Gaussian Cox process model (26). This predicted relative risk surface reveals several small areas of raised risk that are not apparent in Figure 7. The top right-hand panel shows the log of the estimated variance of relative risk. To account for this variation, we produced a plot of the posterior probability that relative risk exceeds 1.1, shown in the bottom panel. This shows that higher rates of incidence appear to be mainly confined to a number of small townships, the largest of which is an area to the north of Toledo and surrounding the Illescas municipality, where there are a number of contiguous cells for which the probability exceeds 0.6.

We acknowledge that this is an illustrative example. In particular, we cannot guarantee the reliability of the estimate of population density used as an offset.

In a discussion of Markov models for spatial data, Wall (2004) investigated properties of the covariance structure implied by the simultaneous and conditional autoregressive models on an irregular lattice. She concluded that the "implied spatial correlation [between cells in these] models does not seem to follow an intuitive or practical scheme" and advises "[using] other ways of modelling lattice data ... should be considered, especially when there is interest in understanding the spatial structure". Our approach is one such. Others, which we discuss in Section 7, include proposals in Best, Ickstadt and Wolpert (2000) and Kelsall and Wakefield (2002).

Our spatially continuous formulation does not entirely rescue us from the trap of the ecological fallacy (Piantadosi, Byar and Green, 1988; Greenland and Morgenstern, 1990). In a spatial context, this refers to the fact that the association between a risk-factor and a health outcome need not be, and usually is not, independent of the spatial scale on which the risk-factor and outcome variables are defined. In our example, we have to accept that treating covariate surfaces as if they were piece-wise constant is a convenient fiction. However, our methodology avoids any necessity to aggregate all covariate and outcome variables to a common set of spatial units, but rather operates at the fine resolution of the computational grid. In effect, this enables us to place a spatially continuous interpretation on any parameters relating to continuously measured components of the model, whether covariates or the latent stochastic process $S(x)$.

6. SPATIO-TEMPORAL LOG-GAUSSIAN COX PROCESSES

6.1 Models

A spatio-temporal LGCP is defined in the obvious way, as a spatio-temporal Poisson point process conditional on the realisation of a stochastic intensity function $\Lambda(x,t) = \exp\{S(x,t)\}$, where $S(\cdot)$ is a Gaussian process. Gneiting and Guttorp (2010b) review the literature on formulating models for spatio-temporal Gaussian processes. They make a useful distinction between physically motivated constructions and more empirical formulations. An example of the former is given in Brown et al. (2000), who propose models based on a physical dispersion process. In discrete time, with δ denoting the time-separation between successive realisations of the spatial field, their model takes the form

$$\begin{aligned} &S(x,t) \\ &\quad = \int h_\delta(u) S(x-u, t-\delta)\, du + Z_\delta(x,t), \end{aligned} \tag{27}$$

where $h_\delta(\cdot)$ is a smoothing kernel and $Z_\delta(\cdot)$ is a noise process, in each case with parameters that depend on the value of δ in such a way as to give a consistent interpretation in the spatio-temporally continuous limit as $\delta \to 0$.

Amongst empirical spatio-temporal covariance models, a basic distinction is between separable and non-separable models. Suppose that $S(x,t)$ is stationary, with variance σ^2 and correlation function $r(u,v) = \mathrm{Corr}\{S(x,t), S(x-u,t-v)\}$. In a *separable* model, $r(u,v) = r_1(u)r_2(v)$, where $r_1(\cdot)$ and $r_2(\cdot)$ are spatial

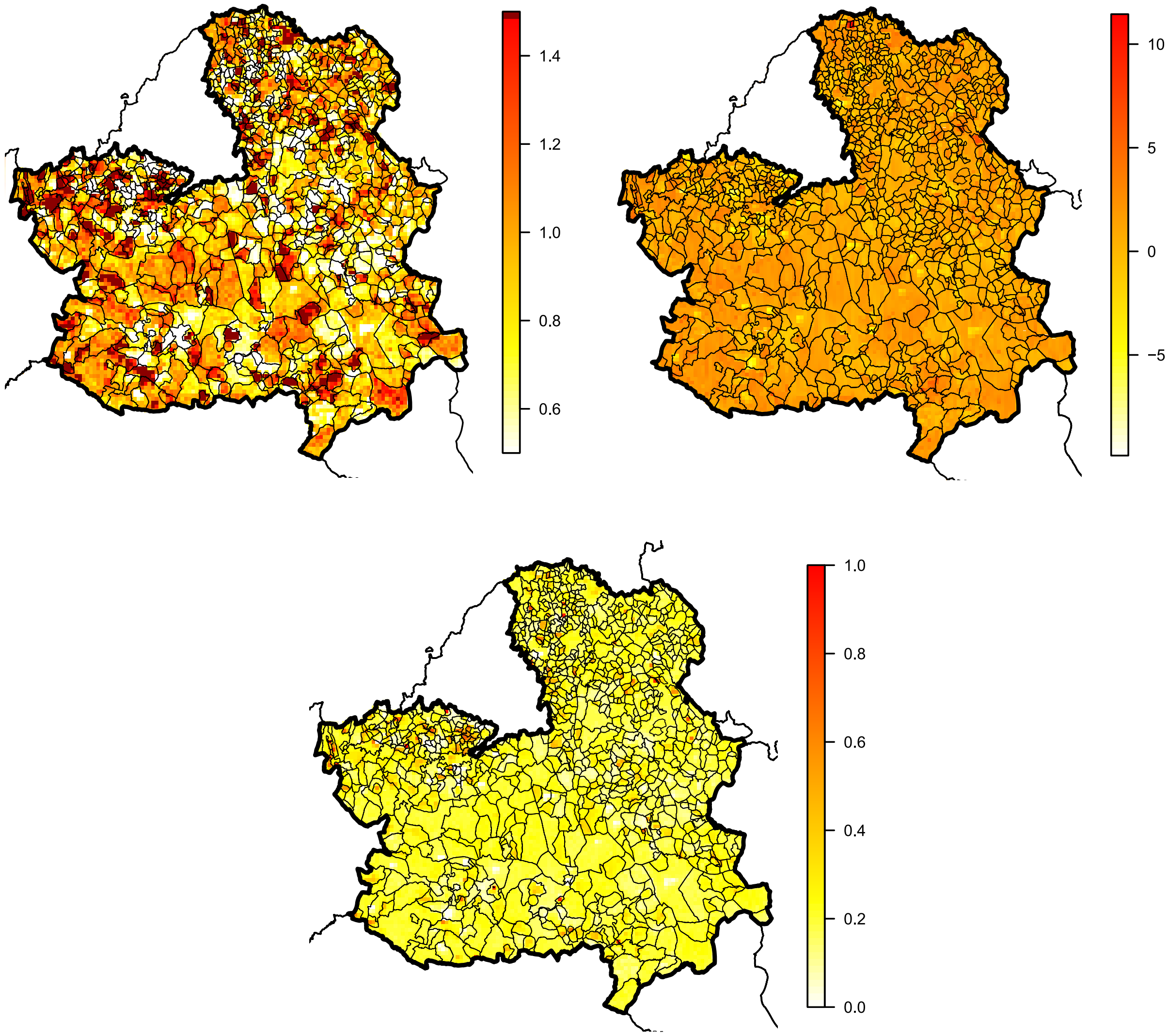

FIG. 9. *Lung Cancer mortality in the Castile-La Mancha Region of Spain. The top left panel shows covariate-adjusted relative risk. The range of values was restricted to lie between* 0.5 *and* 1.5 *to allow comparison with Figure* 7. *Inside the Castile-La Mancha region, cells with mean relative risk greater than* 1.5 *appear dark red and cells with relative risk below* 0.5 *appear white. The top right panel shows the log of the estimated variance of relative risk. The bottom panel shows the predictive probability that the covariate-adjusted relative risk exceeds* 1.1.

and temporal correlation functions. The separability assumption is convenient, not least because any valid specification of $r_1(u)$ and $r_2(v)$ guarantees the validity of $r(u, v)$, but it is not especially natural. Parametric families of nonseparable models are discussed in Cressie and Huang (1999), Gneiting (2002), Ma (2003, 2008) and Rodrigues and Diggle (2010).

As noted by Gneiting and Guttorp (2010b), whilst spatio-temporally continuous processes are, in formal mathematical terms, simply spatially continuous processes with an extra dimension, from a scientific perspective models need to reflect the fundamentally different nature of space and time, and, in particular, time's directional quality. For this reason, in applications where data arise as a set of spatially indexed time-series, a natural way to formulate a spatio-temporal model is as a multivariate time series whose cross-covariance functions are spatially structured. For example, a spatially discrete version of (27) on a finite set of spatial locations $x_i : i = 1, \ldots, n$ and integer times t would be

$$S_{it} = \sum_{j=1}^{n} h_{ij} S_{i,t-1} + Z_{it}, \qquad (28)$$

where the h_{ij} are functions of the corresponding locations, x_i and x_j. For a review of models of this kind, see Gamerman (2010).

6.2 Spatio-Temporal Prediction: Real-Time Monitoring of Gastrointestinal Disease

An early implementation of spatio-temporal log-Gaussian process modelling was used in the AEGISS project (Ascertainment and Enhancement of Gastroenteric Infection Surveillance Statistics, see http://www.maths.lancs.ac.uk/~diggle/Aegiss/day.html%3fyear=2002). The overall aim of the project was to investigate how health-care data routinely collected within the UK's National Health Service (NHS) could be used to spot outbreaks of gastro-intestinal disease. The project is described in detail in Diggle et al. (2003), whilst Diggle, Rowlingson and Su (2005) give details of the spatio-temporal statistical model.

As part of the government's modernisation programme for the NHS, the nonemergency NHS Direct telephone service was launched in the late 1990s, and by 2000 was serving all of England and Wales (http://www.nhsdirect.nhs.uk/About/WhatIsNHSDirect/History). Callers to this 24-hour system were questioned about their problem and advised accordingly. This process reduced calls to an "algorithm code" which was a broad classification of the problem. Basic information on the caller, including age, sex and postal code, was also recorded. Cooper and Chinemana (2004) give a more detailed description of the NHS Direct system. Mark and Shepherd (2004) analyse its impact on the demand for primary care in the UK. Cooper et al. (2003) report a retrospective analysis of 150,000 calls to NHS Direct classified as diarrhoea or vomiting, and concluded that fluctuations in the rate of such calls could be a useful proxy for monitoring the incidence of gastrointestinal illness.

In the AEGISS project, residential postal codes associated with calls classified as relating to diarrhoea or vomiting were converted to grid references using a lookup table. Postal codes at this level are referenced to 100 metre precision, which on the scale of the study area (the county of Hampshire) is effectively continuous. The data then formed a spatio-temporal point pattern.

The daily extraction of data for Hampshire and the location coding was done by the NHS at Southampton. These data were encrypted and sent by email to Lancaster, where the emails were automatically filtered, decrypted and stored. An overnight run of the MALA algorithm described in Brix and Diggle (2001) took the latest data and produced maps of predictive probabilities for the risk exceeding multiples 2, 4 and 8 of the baseline rate.

The specification of the model, based on an exploratory analysis of the data, was a spatio-temporal LGCP with intensity

$$\Lambda(x, t) = \lambda_0(x)\mu_0(t) \exp\{S(x, t)\}.$$

The spatial baseline component, $\lambda_0(x)$, was calculated by a kernel smoothing of the first two years of case locations, whilst the temporal baseline, $\mu_0(t)$, was obtained by fitting a standard Poisson regression model to the counts over time. This regression model included an annual seasonal component, a factor representing the day-of-the-week and a trend term to represent the increasing take-up of the NHS Direct service during the life-time of the project.

The parameters of $S(x, t)$ were then estimated using moment-based methods, as in Brix and Diggle (2001), with a separable correlation structure. Uncertainty in these parameter estimates was considered to have a minimal effect on the predictive distribution of $S(x, t)$ because parameter estimates are informed by all of the data, whereas prediction of $S(x, t)$ given the model parameters benefits only from data points that lie close to (x, t), that is, within the range of the spatio-temporal correlation.

Plug-in predictive inference was then performed using the MALA algorithm on each new set of data arriving overnight. Instead of storing the outputs from each of 10,000 iterations, only a count of where $S(x, t)$ exceeded a threshold that corresponded to 2, 4 or 8 times the baseline risk was retained. This range of thresholds was chosen in consultation with clinicians; a doubling of risk was considered of possible interest, whilst an eightfold increase was considered potentially serious. These exceedence counts were then converted into exceedence probabilities.

Presentation of these exceedence maps was an important aspect of the AEGISS project. At the time, there were few implementations of maps on the internet—UMN MapServer was released as open source in 1997 and the Google Maps service started in 2005. A simpler approach was used where static images of the exceedence probabilities were generated by R's graphics system. Regions where the exceedence probability was higher than 0.9 were outlined with a box and displayed in a zoomed-in version below the main graphic. Other page controls enabled the user to select the threshold value as 2, 4 or 8, and to select a day or month. A traffic light system of green, amber and red warnings dependent on the severity of exceedence threshold crossings was developed for rapid assessment of conditions on any particular day. The left-hand panel of Figure 10 shows a day where two clusters

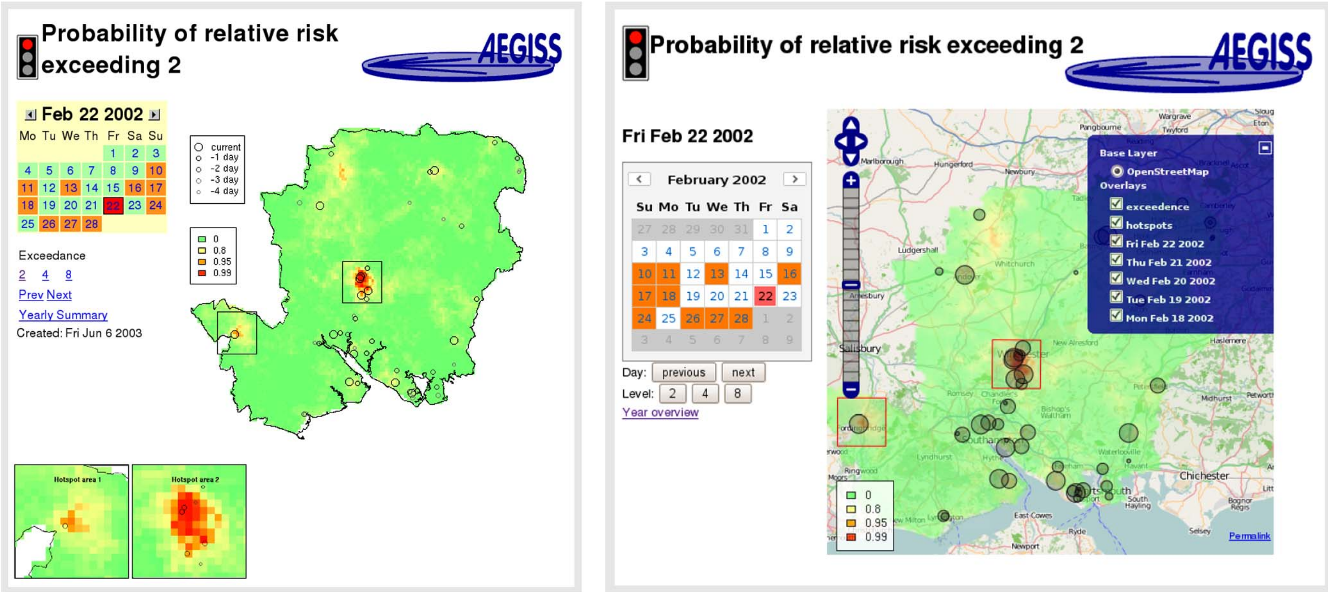

FIG. 10. *AEGISS web page design. Left-hand panel shows the original design, right-hand panel a modern redesign.*

of grid cells show high predictive probability of at least a doubling of risk relative to baseline.

With modern web-based technologies the user interface could be constructed as a dynamic web-mapping system that would allow the user freely to navigate the study region. Layers of information, such as cases or exceedence probability maps, can then be selected by the user as overlays. The right-hand panel of Figure 10 shows the same day as the left-hand panel, but uses the OpenLayers (http://www.openlayers.org) web-mapping toolkit to superimpose the cases and risk surface on a base map composed of data from OpenStreetMap (http://www.openstreetmap.org). This also shows the layer selector menu for further customisation.

Increases in computing power and algorithmic advances mean that longer MCMC runs can be performed overnight or on finer spatial resolutions. However, increasing ethical concerns over data use and patient confidentiality mean that finely resolved spatio-temporal data are becoming harder to obtain. Recent changes in the organisation of the NHS 24-hour telephone helpline has meant that several providers will now be responsible for regional services contributing to a new system, NHS111 (http://www.nhs.uk/111). AEGISS was originally conceived as a pilot project that could be rolled out to all of the UK, but obtaining data from all the new providers and dealing with possible systematic differences between them in order to perform a statistically rigorous analysis is now more challenging. The future of health surveillance systems may lie in the use of multivariate spatio-temporal models to combine information from multiple data streams including nontraditional proxies for health outcomes, such as nonprescription medicine sales, counts of key words and phrases used in search engine queries, and text-mining of social media sites.

7. DATA SYNTHESIS: INTEGRATED ANALYSIS OF EXPOSURE AND HEALTH OUTCOME DATA AT MULTIPLE SPATIAL SCALES

The ubiquitous problem of dealing with exposure and health outcome data recorded at disparate spatial scales is known to geographers as the "modifiable areal unit problem." See, for example, the reviews by Gotway and Young (2002) and Dark and Bram (2007). In the statistical literature, a more common term is "spatial misalignment." See, for example, Gelfand (2010). Several authors have considered special cases of this problem in an epidemiological setting. Mugglin, Carlin and Gelfand (2000) deal with data in the form of disease counts on a partition of the region of interest, A, into a discrete set of subregions, A_i, together with covariate information on a different partition, B_i, say. Their solution is based on creating a single, finer partition that includes all nonzero intersections $A_i \cap B_j$. Best, Ickstadt and Wolpert (2000) also consider count data on a discrete partition of A, but assume that covariate information on a risk factor of interest is available throughout A. They consider count data to be derived from an underlying Cox process whose intensity varies in a spatially continuous manner through the combination of a covariate effect and a latent stochastic process modelled as a kernel-smoothed gamma random field. They then derive the distribution of the observed counts by spatial integration over the A_i. Kelsall and Wakefield (2002) take a similar approach, but using a log-Gaussian latent stochastic process rather than a gamma random field. The technical and computational issues that arise when handling spatial integrals of stochastic processes can be simplified by using low-rank models, such as the class of Gaussian predictive process models proposed by Banerjee et al. (2008) and further developed by Finley et al. (2009). Gelfand (2012) gives a useful summary of this and related work.

All of these approaches can be subsumed within a single modelling framework for multiple exposures and disease risk by considering these as a set of spatially continuous processes, irrespective of the spatial resolution at which data elements are recorded. For example, a model for the spatial association between disease risk, $R(x)$, and m exposures $T_k(x): k = 1, \ldots, m$ can be obtained by treating individual case-locations as an LGCP with intensity

$$R(x) = \exp\left\{\alpha + \sum_{k=1}^{p} \beta_k T_k(x) + S(x)\right\}, \tag{29}$$

where $S(x)$ denotes stochastic variation in risk that is not captured by the p covariate processes $T_k(x)$. The inferential algorithms associated with model (29) would then depend on the structure of the available data.

Suppose, for example, that health outcome data are available in the form of area-level counts, $Y_i : i = 1, \ldots, n$, in subregions A_i, whilst exposure data are obtained as collections of unbiased estimates, U_{ik}, of the $T_k(x)$ at corresponding locations $x_{ik} : i = 1, \ldots, m_k$. Suppose further that the U_{ik} are conditionally independent, with $U_{ik}|T_k(\cdot) \sim \mathrm{N}(T_k(x_{ik}), \tau_k^2)$, the processes $T_k(\cdot)$ are jointly Gaussian and the process $S(\cdot)$ is also Gaussian and independent of the $T_k(\cdot)$. A possible inferential goal is to evaluate the predictive distribution of the risk surface $R(\cdot)$ given the data $Y_i : i = 1, \ldots, m$ and $U_{ik} : i = 1, \ldots, m_k; k = 1, \ldots, p$. In an obvious shorthand, and temporarily ignoring the issue of parameter estimation, the required predictive distribution is $[S, T|U, Y]$. The joint distribution of S, T, U and Y factorises as

$$[S, T, U, Y] = [S][T][U|T][Y|S, T], \tag{30}$$

where $[S]$ and $[T]$ are multivariate Gaussian densities, $[U|T]$ is a product of univariate Gaussian densities, and $[Y|S, T]$ is a product of Poisson probability distributions with means

$$\mu_i = \int_{A_i} R(x)\, dx.$$

Sampling from the required predictive distributions can then proceed using a suitable MCMC algorithm. For Bayesian parameter estimation, we would augment (30) by a suitable joint prior for the model parameters before designing the MCMC algorithm.

A specific example of data synthesis concerns an ongoing leptospirosis cohort study in a poor community within the city of Salvador, Brazil. Leptospirosis is considered to be the most widespread of the zoonotic diseases. This is due to the large number of people worldwide, but especially in poor communities, who live in close proximity to wild and domestic mammals that serve as reservoirs of infection and shed the agent in their urine. The major mode of transmission is contact with contaminated water or soil (Levett, 2001; Bharti et al., 2003; McBride et al., 2005). In the majority of cases infection leads to an asymptomatic or mild, self-limiting febrile illness. However, severe cases can lead to potentially fatal acute renal failure and pulmonary haemorrhage syndrome. Leptospirosis is traditionally associated with rural-based subsistence farming communities, but rapid urbanization and widening social inequality have led to the dramatic growth of urban slums, where the lack of basic sanitation favours rat-borne transmission (Ko et al., 1999; Johnson et al., 2004).

The goals of the cohort study are to investigate the combined effects of social and physical environmental factors on disease risk, and to map the unexplained spatio-temporal variation in incidence. In the study, approximately 1700 subjects $i = 1, \ldots, n$ at residential locations x_i provide blood-samples on recruitment and at subsequent times t_{ij} approximately 6, 12, 18 and 24 months later. At each post-recruitment visit, sero-conversion is defined as a change from zero to positive, or at least a fourfold increase in concentration. The resulting data consist of binary responses, $Y_{ij} = 0/1 : j = 1, 2, 3, 4$ (sero-conversion no/yes), together with a mix of time-constant and time-varying risk-factors, r_{ij}.

A conventional analysis might treat the data from each subject as a time-sequence of binary responses with associated explanatory variables. Widely used methods for data of this kind include generalised estimating equations (Liang and Zeger, 1986) and generalised linear mixed models (Breslow and Clayton, 1993). An analysis more in keeping with the philosophy of the current paper would proceed as follows.

Let a_i and $b_i(t)$ denote time-constant and time-varying explanatory variables associated with subject i, and t_{ij} the times at which blood samples are taken, setting $t_{i0} = 0$ for all i. Note that explanatory variables can be of two distinct kinds: characteristics of an individual subject, for example, their age; and characteristics of a subject's place of residence, for example, its proximity to an open-sewer. In principle, the latter can be indexed by a spatially continuous location, hence, $a_i = A(x_i)$ and $b_i(t) = B(x_i, t)$. A response $Y_{ij} = 1$ indicates that at least one infection event has occurred in the time-interval $(t_{i,j-1}, t_{ij})$. A model for each subject's risk of infection then requires the specification of a set of person-specific hazard functions, $\Lambda_i(t)$. A model that allows for unmeasured risk factors would be a set of LGCPs, one for each subject, with respective stochastic intensities,

$$\Lambda_i(t) = \exp\{a_i'\alpha + b_i(t_{ij})'\beta + U_i + S(x_i, t)\}, \tag{31}$$

where the U_i are mutually independent $\mathrm{N}(0, \nu^2)$ and $S(x, t)$ is a spatio-temporally continuous Gaussian process. It follows that

$$\begin{aligned} &\mathrm{P}\{Y_{it} = 1 | \Lambda_i(\cdot)\} \\ &\quad = 1 - \exp\left\{-\int_{t_{i,j-1}}^{t_{ij}} \Lambda_i(u)\, du\right\}. \end{aligned} \tag{32}$$

In practice, values of $a(x)$ and $b(x, t)$ may only be observed incompletely, either at a finite number of locations or as small-area averages. For notational convenience, we consider only a single, incompletely observed spatio-temporal covariate whose measured values, $b_k : k = 1, \dots, m$, we model as

$$b_k = B(x_k, t_k) + Z_k, \tag{33}$$

where $B(x, t)$ is a spatio-temporal Gaussian process and the Z_k are mutually independent $\mathrm{N}(0, \tau^2)$ measurement errors. Then, (31) becomes

$$\Lambda_i(t) = \exp\{B(x_i, t_{ij})'\beta + U_i + S(x_i, t)\}. \tag{34}$$

Inference for the model defined by (32), (33) and (34), based on data $\{y_{ij} : j = 1, \dots, 4; i = 1, \dots, n\}$ and $b = \{b_k : k = 1, \dots, m\}$, would require further development of MCMC algorithms of the kind described in Section 4.

8. DISCUSSION

In this paper we have argued that the LGCP provides a useful class of models, not only for point process data but also for any problem involving prediction of an incompletely observed spatial or spatio-temporal process, irrespective of data format. Developments in statistical computation have made the combination of likelihood-based, classical or Bayesian parameter estimation and probabilistic prediction feasible for relatively large data sets, including real-time updating of spatio-temporal predictions.

In each of our applications, the focus has been on prediction of the spatial or spatio-temporal variation in a response surface, rather than on estimation of model parameters. In problems of this kind, where parameters are not of direct interest but rather are a means to an end, Bayesian prediction in conjunction with diffuse priors is an attractive strategy, as its predictions naturally accommodate the effect of parameter uncertainty. Model-based predictions are essentially nonparametric smoothers, but embedded within a probabilistic framework. This encourages the user to present results in a way that emphasises, rather than hides, their inherent imprecision.

In many public health settings, identifying where and when a particular phenomenon, such as disease incidence, is likely to have exceeded an agreed intervention threshold is more useful than quoting either a point estimate and its standard error or the statistical significance of departure from a benchmark.

The log-linear formulation is convenient because of the tractable moment properties of the log-Gaussian distribution. It also gives the model a natural interpretation as a multiplicative decomposition of the overall intensity into deterministic and stochastic components. However, it can lead to very highly skewed marginal distributions, with large patches of near-zero intensity interspersed with sharp peaks. Within the Monte Carlo inferential framework, there is no reason why other, less severe transformations from $\mathbb{R}$ to $\mathbb{R}^+$ should not be used.

Two areas of current methodological research are the formulation of models and methods for principled analysis of multiple data streams that include data of variable quality from nontraditional sources, and the further development of robust computational algorithms that can deliver reliable inferences for problems of ever-increasing complexity.

Our general approach reflects a continuing trend in applied statistics since the 1980s. The explosion in the development of computationally intensive methods and associated complex stochastic models has encouraged a move away from a methods-based classification of the statistics discipline and towards a multidisciplinary, problem-based focus in which statistical method (singular) is thoroughly embedded within scientific method.

ACKNOWLEDGEMENTS

We thank the Department of Environmental and Cancer Epidemiology in the National Center For Epidemiology (Spain) for providing aggregated data from the Castile-La Mancha region for permission to use the Spanish lung cancer data.

The leptospirosis study described in Section 7 is funded by a USA National Science Foundation grant, with Principal Investigator Professor Albert Ko (Yale University School of Public Health). This work was supported by the UK Medical Research Council (Grant number G0902153).

SUPPLEMENTARY MATERIAL

Supplementary materials for "Spatial and spatio-temporal log-Gaussian Cox processes: Extending the geostatistical paradigm" (DOI: 10.1214/13-STS441SUPP; .pdf). This material contains mixing, convergence and inferential diagnostics for all of the examples in the main article and is also available from http://www.lancs.ac.uk/staff/taylorb1/statsciappendix.pdf.

REFERENCES

ANDRIEU, C. and THOMS, J. (2008). A tutorial on adaptive MCMC. *Stat. Comput.* **18** 343–373. MR2461882

BADDELEY, A. J., MØLLER, J. and WAAGEPETERSEN, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Stat. Neerl.* **54** 329–350. MR1804002

BANERJEE, S., GELFAND, A. E., FINLEY, A. O. and SANG, H. (2008). Gaussian predictive process models for large spatial data sets. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **70** 825–848. MR2523906

BARTLETT, M. S. (1964). The spectral analysis of two-dimensional point processes. *Biometrika* **51** 299–311. MR0175254

BARTLETT, M. S. (1975). *The Statistical Analysis of Spatial Pattern*. Chapman & Hall, London. MR0402886

BERMAN, M. and DIGGLE, P. (1989). Estimating weighted integrals of the second-order intensity of a spatial point process. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **51** 81–92. MR0984995

BESAG, J., YORK, J. and MOLLIÉ, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.* **43** 1–59. MR1105822

BEST, N. G., ICKSTADT, K. and WOLPERT, R. L. (2000). Spatial Poisson regression for health and exposure data measured at disparate resolutions. *J. Amer. Statist. Assoc.* **95** 1076–1088. MR1821716

BHARTI, A. R., NALLY, J. E., RICALDI, J. N., MATTHIAS, M. A., DIAZ, M. M., LOVETT, M. A., LEVETT, P. N., GILMAN, R. H., WILLIG, M. R., GOTUZZO, E. and VINETZ, J. M. (2003). Leptospirosis: A zoonotic disease of global importance. *Lancet. Infect. Dis.* **3** 757–771.

BRESLOW, N. E. and CLAYTON, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.* **88** 9–25.

BRIX, A. and DIGGLE, P. J. (2001). Spatiotemporal prediction for log-Gaussian Cox processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63** 823–841. MR1872069

BRIX, A. and DIGGLE, P. J. (2003). Corrigendum: Spatio-temporal prediction for log-Gaussian Cox processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65** 946.

BROWN, P. E., KÅRESEN, K. F., ROBERTS, G. O. and TONELLATO, S. (2000). Blur-generated non-separable space–time models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **62** 847–860. MR1796297

COOPER, D. and CHINEMANA, F. (2004). NHS direct derived data: An exciting new opportunity or an epidemiological headache? *J. Public Health* (*Oxf.*) **26** 158–160.

COOPER, D. L., SMITH, G. E., O'BRIEN, S. J., HOLLYOAK, V. A. and BAKER, M. (2003). What can analysis of calls to NHS direct tell us about the epidemiology of gastrointestinal infections in the community? *J. Infect.* **46** 101–105.

COX, D. R. (1955). Some statistical methods connected with series of events. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **17** 129–157; discussion, 157–164. MR0092301

CRESSIE, N. A. C. (1991). *Statistics for Spatial Data*. Wiley, New York. MR1127423

CRESSIE, N. and HUANG, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *J. Amer. Statist. Assoc.* **94** 1330–1340. MR1731494

DARK, S. J. and BRAM, D. (2007). The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography* **31** 471–479.

DIGGLE, P. J. and RIBEIRO, P. J. JR. (2007). *Model-Based Geostatistics*. Springer, New York. MR2293378

DIGGLE, P., ROWLINGSON, B. and SU, T.-L. (2005). Point process methodology for on-line spatio-temporal disease surveillance. *Environmetrics* **16** 423–434. MR2147534

DIGGLE, P. J., ZHENG, P. and DURR, P. (2005). Non-parametric estimation of spatial segregation in a multivariate point process. *Applied Statistics* **54** 645–658.

DIGGLE, P. J., KNORR-HELD, L., ROWLINGSON, B., SU, T., HAWTIN, P. and BRYANT, T. (2003). Towards on-line spatial surveillance. In *Monitoring the Health of Populations*: *Statistical Methods for Public Health Surveillance* (R. Brookmeyer and D. Stroup, eds.). Oxford Univ. Press, Oxford.

DIGGLE, P. J., MORAGA, P., ROWLINGSON, B. and TAYLOR, B. M. (2013). Supplement to "Spatial and spatio-temporal log-Gaussian Cox processes: Extending the geostatistical paradigm." DOI:10.1214/13-STS441SUPP.

DONNELLY, C. A., WOODROFFE, R., COX, D. R., BOURNE, F. J., CHEESMAN, C. L., CLIFTON-HADLEY, R. S., WEI, G., GETTINBY, G., GILKS, P., JENKINS, H., JOHNSTON, W. T., LE FEVRE, A. M., MCINERY, J. P. and MORRISON, W. I. (2006). Positive and negative effects of widespread badger culling on tuberculosis in cattle. *Nature* **485** 843–846.

FINLEY, A. O., SANG, H., BANERJEE, S. and GELFAND, A. E. (2009). Improving the performance of predictive process modeling for large datasets. *Comput. Statist. Data Anal.* **53** 2873–2884. MR2667597

FRIGO, M. and JOHNSON, S. G. (2011). FFTW fastest Fourier transform in the west. Available at http://www.fftw.org/.

GAMERMAN, D. (2010). Dynamic spatial models including spatial time series. In *Handbook of Spatial Statistics* (A. E. Gelfand, P. J. Diggle, M. Fuentes and P. Guttorp, eds.) 437–448. CRC Press, Boca Raton, FL. MR2730959

GAMERMAN, D. and LOPES, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd ed. Chapman & Hall/CRC, Boca Raton, FL. MR2260716

GELFAND, A. E. (2010). Misaligned spatial data: The change of support problem. In *Handbook of Spatial Statistics* (A. E. Gelfand, P. J. Diggle, M. Fuentes and P. Guttorp, eds.) 517–539. CRC Press, Boca Raton, FL. MR2730964

GELFAND, A. E. (2012). Hierarchical modelling for spatial data problems. *Spatial Statistics* **1** 30–39.

GELFAND, A. E., DIGGLE, P. J., FUENTES, M. and GUTTORP, P., eds. (2010). *Handbook of Spatial Statistics*. CRC Press, Boca Raton, FL. MR2761512

GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern. Anal. Mach. Intell.* **6** 721–741.

GERRARD, D. J. (1969). Competition Quotient: A New Measure of the Competition Affecting Individual Forest Trees. *Research Bulletin* **20**. Agricultural Experiment Station, Michigan State Univ., East Lansing, MI.

GEYER, C. (1999). Likelihood inference for spatial point processes: Likelihood and computation. In *Stochastic Geometry* (*Toulouse*, 1996), (O. E. Barndorff-Nielsen, W. S. Kendall and M. N. M. van Lieshout, eds.). *Monogr. Statist. Appl. Probab.* **80** 79–140. Chapman & Hall/CRC, Boca Raton, FL. MR1673118

GILKS, W., RICHARDSON, S. and SPIEGELHALTER, D. (1995). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

GIROLAMI, M. and CALDERHEAD, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73** 123–214. MR2814492

GNEITING, T. (2002). Nonseparable, stationary covariance functions for space–time data. *J. Amer. Statist. Assoc.* **97** 590–600. MR1941475

GNEITING, T. and GUTTORP, P. (2010a). Continuous parameter stochastic process theory. In *Handbook of Spatial Statistics* (A. E. Gelfand, P. J. Diggle, M. Fuentes and P. Guttorp, eds.) 17–28. CRC Press, Boca Raton, FL. MR2730952

GNEITING, T. and GUTTORP, P. (2010b). Continuous parameter spatio-temporal processes. In *Handbook of Spatial Statistics* (A. E. Gelfand, P. J. Diggle, M. Fuentes and P. Guttorp, eds.) 427–436. CRC Press, Boca Raton, FL. MR2730958

GOTWAY, C. A. and YOUNG, L. J. (2002). Combining incompatible spatial data. *J. Amer. Statist. Assoc.* **97** 632–648. MR1951636

GREENLAND, S. and MORGENSTERN, H. (1990). Ecological bias, confounding and effect modification. *International Journal of Epidemiology* **18** 269–274.

HARAN, M. and TIERNEY, L. (2012). On automating Markov chain Monte Carlo for a class of spatial models. Available at http://arxiv.org/abs/1205.0499.

HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109.

JOHNSON, M. A., SMITH, H., JOSEPH, P., GILMAN, R. H., BAUTISTA, C. T., CAMPOS, K. J., CESPEDES, M., KLATSKY, P., VIDAL, C., TERRY, H., CALDERON, M. M., CORAL, C., CABRERA, L., PARMAR, P. S. and VINETZ, J. M. (2004). Environmental exposure and leptospirosis, Peru. *Emerging Infectious Diseases* **10** 1016–1022.

KELSALL, J. and WAKEFIELD, J. (2002). Modeling spatial variation in disease risk: A geostatistical approach. *J. Amer. Statist. Assoc.* **97** 692–701. MR1941405

KO, A. I., REIS, M. G., DOURADO, C. M. R., JOHNSON, W. D. JR. and RILEY, L. W. (1999). Urban epidemic of severe leptospirosis in Brazil. Salvador Leptospirosis Study Group. *Lancet* **354** 820–825.

LEVETT, P. N. (2001). Leptospirosis. *Clininical Microbiology Reviews* **14** 296–326.

LI, Y., BROWN, P. E., GESINK, D. C. and RUE, H. (2012). Log Gaussian Cox processes and spatially aggregated disease incidence data. *Stat. Methods Med. Res.* **21** 479–507.

LIANG, K. Y. and ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22. MR0836430

LINDGREN, F., RUE, H. and LINDSTRÖM, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73** 423–498. MR2853727

LÓPEZ-ABENTE, G., RAMIS, R., POLLÁN, M., ARAGONÉS, N., PÉREZ-GÓMEZ, B., GÓMEZ-BARROSO, D., CARRASCO, J. M., LOPE, V., GARCÍA-PÉREZ, J., BOLDO, E. and GARCÍA-MENDIZÁBAL, M. J. (2006). ATLAS municipal de mortalidad por cáncer en España, 1989–1998. Instituto de Salud Carlos III, Madrid.

MA, C. (2003). Families of spatio-temporal stationary covariance models. *J. Statist. Plann. Inference* **116** 489–501. MR2000096

MA, C. (2008). Recent developments on the construction of spatio-temporal covariance models. *Stoch. Environ. Res. Risk Assess.* **22** 39–47. MR2418410

MARK, A. L. and SHEPHERD, D. H. (2004). NHS Direct: Managing demand for primary care? *International Journal of Health Planning and Management* **19** 79–91.

MATÉRN, B. (1960). *Spatial Variation*. Meddelanden fran Statens Skogsforskningsinstitut, Stockholm. Band 49, number 5. MR0169346

MCBRIDE, A. J., ATHANAZIO, D. A., REIS, M. G. and KO, A. I. (2005). Leptospirosis. *Current Opinions in Infectious Diseases* **18** 376–386.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* **21** 1087–1092.

MØLLER, J., SYVERSVEEN, A. R. and WAAGEPETERSEN, R. P. (1998). Log Gaussian Cox processes. *Scand. J. Stat.* **25** 451–482. MR1650019

MUGGLIN, A. S., CARLIN, B. P. and GELFAND, A. E. (2000). Fully model-based approaches for spatially misaligned data. *J. Amer. Statist. Assoc.* **95** 877–887.

PIANTADOSI, S., BYAR, D. P. and GREEN, S. B. (1988). The ecological fallacy. *Am. J. Epidemiol.* **127** 893–904.

R CORE TEAM. (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

RIPLEY, B. D. (1976). The second-order analysis of stationary point processes. *J. Appl. Probab.* **13** 255–266. MR0402918

RIPLEY, B. D. (1977). Modelling spatial patterns. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **39** 172–212. MR0488279

ROBERTS, G. O. and ROSENTHAL, J. S. (2001). Optimal scaling for various Metropolis–Hastings algorithms. *Statist. Sci.* **16** 351–367. MR1888450

ROBERTS, G. O. and ROSENTHAL, J. S. (2007). Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *J. Appl. Probab.* **44** 458–475. MR2340211

ROBERTS, G. O. and TWEEDIE, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. MR1440273

RODRIGUES, A. and DIGGLE, P. J. (2010). A class of convolution-based models for spatio-temporal processes with non-separable covariance structure. *Scand. J. Stat.* **37** 553–567. MR2779636

RUE, H. and HELD, L. (2005). *Gaussian Markov Random Fields: Theory and Applications. Monographs on Statistics and Applied Probability* **104**. Chapman & Hall/CRC, Boca Raton, FL. MR2130347

RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** 319–392. MR2649602

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London. MR0848134

SPIEGELHALTER, D. J., THOMAS, A. and BEST, N. G. (1999). WinBUGS Version 1.2 User Manual.

TAYLOR, B. M. and DIGGLE, P. J. (2013a). INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. *J. Stat. Comput. Simul.* To appear. Preprint available at http://arxiv.org/abs/1202.1738.

TAYLOR, B. M. and DIGGLE, P. J. (2013b). Corrigendum: Spatiotemporal prediction for log-Gaussian Cox processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **75** 601–602. MR3065481

TAYLOR, B. M., DAVIES, T. M., ROWLINGSON, B. S. and DIGGLE, P. J. (2013). lgcp: Inference with spatial and spatio-temporal log-Gaussian Cox processes in R. *Journal of Statistical Software* **52** Issue 4.

WALL, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *J. Statist. Plann. Inference* **121** 311–324. MR2038824

WOOD, A. T. A. and CHAN, G. (1994). Simulation of stationary Gaussian processes in $[0, 1]^d$. *J. Comput. Graph. Statist.* **3** 409–432. MR1323050

WOODROFFE, R., DONNELLY, C. A., JOHNSTON, W. T., BOURNE, F. J., CHEESMAN, C. L., CLIFTON-HADLEY, R. S., COX, D. R., GETTINBY, G., HEWINSON, R. G., LE FEVRE, A. M., MCINERY, J. P. and MORRISON, W. I. (2005). Spatial association of *Mycobacterium bovis* infection in cattle and badgers *Meles meles*. *Journal of Applied Ecology* **42** 852–862.

ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99** 250–261. MR2054303